NEXT-GENERATION SEQUENCING AND MOTIF GRAFTING APPLICATIONS IN SYNTHETIC ANTIBODY DISCOVERY

A Thesis Submitted to the

College of Graduate and Postdoctoral Studies

In Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy

In the Department of Biochemistry

University of Saskatchewan

Saskatoon

By

Bharathikumar Vellalore Maruthachalam

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Prof. C. Ronald Geyer
Department of Pathology
4D01.11 Health Sciences Building
107 Wiggins Road
University of Saskatchewan
Saskatoon, Saskatchewan, S7N 5E5

ABSTRACT

The overall objective of this PhD project was to develop and validate methods for advancing the applications of two techniques, next-generation sequencing (NGS) and motif grafting, in synthetic antibody discovery. In the first part of this project, we developed an NGS-assisted antibody discovery platform by integrating phage-displayed single-framework synthetic antigen-binding fragment (Fab) libraries with Ion Torrent sequencing. We constructed a new single-framework synthetic Fab library containing 8.5 billion unique Fab clones, and validated its functionality by generating high affinity Fabs against Notch and Jagged receptors. We developed a rapid and simple method to link and sequence all diversified complementarity-determining regions (CDRs) in phage Fab pools without losing the CDR pairing information. We identified and reconstructed low-frequency rare Fab clones from NGS information in a reliable and high-throughput manner. In some cases, reconstructed rare clones (frequency ~0.1%) showed higher affinity and better specificity than high-frequency top clones isolated by Sanger sequencing, highlighting the importance of NGS in synthetic antibody discovery. In the second part of this project, we employed motif grafting to semi-rationally design phage-displayed synthetic Fab libraries that are biased towards interacting with a specific site on a receptor. We used structural information on the epidermal growth factor receptor (EGFR) homo-dimerization interaction to design a structure-guided Fab library that was biased towards interacting with domain II of EGFR. We used this structure-guided Fab library to obtain Fabs against the EGFR extracellular domain. For comparison, we used a naïve synthetic Fab library to generate an anti-EGFR Fab whose binding overlapped with the Fab isolated from the structure-guided Fab library. Both Fabs possessed low-nM binding values for recombinant and cell-surface EGFR and inhibited EGF-mediated EGFR activation. Epitope mapping showed that domain II is partially responsible for the interaction of Fabs with EGFR. Further, both Fabs target unique epitopes that are different from previously validated epitopes on EGFR. In total, this PhD project resulted in novel methods for discovering synthetic antibodies using NGS and motif grafting techniques, three functional Fab libraries and numerous high-affinity Fabs against Notch, Jagged and EGF receptors.

ACKNOWLEDGEMENTS

I have been very fortunate to have Prof. Ron Geyer as my mentor. I am deeply grateful to him for providing me a dynamic and challenging project, for allowing me to pursue some of my ideas, and for having confidence in my work. His kindness, guidance and support have been integral in the completion of this work.

I owe an immense debt of gratitude to my graduate committee members, Drs. Jeremy Lee, Stanley Moore, Anthony Kusalik, Yu Luo and Scot Leary, for their invaluable contribution towards shaping my scientific career.

I would like to thank Dr. John DeCoteau for providing facilities and support to conduct next-generation sequencing and cell biological studies.

I extend my sincere thanks to Drs. Sachdev Sidhu and Shane Miersch at the University of Toronto for the successful research collaboration.

I would like to thank the following lab members: Wayne Hill, Lindsay Pelzer, Karen Mochoruk, Landon Pastushok, Kris Barreto, Daniel Hogan, Scot Adams, Ruiyi Guo and Ashley Sutherland, for all the help provided to me in various forms.

I would like to thank Western Economic Diversification Canada for funding this work. I would also like to thank the Department of Biochemistry, College of Medicine, College of Graduate Studies and Research, Canadian Institutes of Health Research- Training in Health Research Using Synchrotron Techniques (CIHR-THRUST), Canadian Light Source, University of Saskatchewan, Government of Saskatchewan, and Canadian Cancer Society for offering various Graduate Scholarships and Travel Awards.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABS	Absorbance
ACVR2B	Activin receptor type-2B
ADAM	A disintegrin and metalloproteinase
ADC	Antibody-drug conjugate
ADCC	Antibody-mediated cellular cytotoxicity
APC	Antigen-presenting cell
ATP	Adenosine triphosphate
B-LyS	B Lymphocyte Stimulator
BRAF	Murine sarcoma viral oncogene homolog B
BSA	Bovine serum albumin
BT-Fab	Biotinylated antigen-binding fragment
CAR	Chimeric antigen receptor
CAT	Cambridge antibody technology
CBP/p300	CREB-binding protein/ E1A binding protein p300
CCC-dsDNA	Covalently-closed, circular, double-stranded deoxyribonucleic acid
CD20/22	Cluster of differentiation 20/22
CDC	Complement-dependent cytotoxicity
cDNA	Complementary deoxyribonucleic acid
CDR	Complementarity-determining region
CFU	Colony-forming units
C_H	Constant heavy domain
C_L	Constant light domain
CSL	C-promoter-binding factor
CTX	Cetuximab
D28 EGFR	ΔC240-C267 Epidermal growth factor receptor
DLL-1/3/4	Delta like canonical Notch ligand-1/3/4
DM EGFR	Y251A + R285S Epidermal growth factor receptor
DMEM	Dulbecco's modified Eagle's medium
DNA	Deoxyribonucleic acid

dNTP	Deoxy-nucleoside triphosphate
dsDNA	Double-stranded deoxyribonucleic acid
DSL	Delta/Serrate/LAG-2
DTT	Dithiothreitol
dU-ssDNA	Deoxy-uridine single-stranded DNA
E. coli	*Escherichia coli*
EC_{50}	Half maximal effective concentration
ECD	Extracellular domain
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
ELISA	Enzyme-linked immunosorbent assay
EMA	European Medicines Agency
EMEM/EBSS	Eagle's minimum essential media with Earle's balanced salt solution
Fab	Antigen-binding fragment
FBS	Fetal bovine serum
Fc	Crystallizable fragment
FcRn	Neo-natal Fc receptors
FcγR	Fc gamma receptors
FDA	Food and Drug Administration
FGF23	Fibroblast Growth Factor 23
FRM	Framework
Fv	Variable fragment
GFP	Green fluorescent protein
GP130	Glycoprotein 130
HER	Human epidermal growth factor receptor
HIV	Human immunodeficiency virus
hr	Hour
HRP	Horseradish peroxidase
HuCAL	Human Combinatorial antibody library
ICN	Intracellular Notch
Ig	Immunoglobulin

IGF1R	Insulin-like growth factor 1 receptor
IL	Interleukin
IL-R	Interleukin receptor
ISPs	Ion sphere particles
ITAM	Immune-receptor tyrosine-based activation motifs
K_D	Equilibrium dissociation constant
K_{OFF}	Dissociation rate constant
K_{ON}	Association rate constant
KRAS	V-Ki-ras2 Kirsten rat sarcoma viral oncogene
mAbs	Monoclonal antibodies
MAC	Membrane attack complex
MAM	Mastermind co-activator
MAPK	Mitogen activated protein kinase
MBP	Maltose-binding protein
MERTK	MER proto-oncogene tyrosine kinase receptor
MHC	Major histocompatibility complex
min	Minute
MNNL	Module at the N-terminus of Notch ligands
NEC	Notch extracellular
NGS	Next-generation sequencing
NK cells	Natural killer cells
OD	Optical density
PB	Phosphate buffered saline with 0.5% BSA
PBL	Peripheral blood lymphocytes
PBS	Phosphate buffered saline
PBT	Phosphate buffered saline with 0.5% BSA and 0.05% Tween 20
PCR	Polymerase chain reaction
PD-L1	Programmed death-ligand 1
PDGFRα	Platelet-derived growth factor receptor α
PE	Phycoerythrin
PEG	Polyethylene glycol

PFU	Plaque-forming units
PGM	Personal Genome Machine
PI3K	Phosphoinositide 3-kinase
PMB	Panitumumab
PT	Phosphate buffered saline with 0.05% Tween 20
PTEN	Phosphatase and tensin homolog
PVDF	Polyvinylidene difluoride
rpm	Rotations per minute
RPMI	Roswell Park Memorial Institute medium
RT	Room temperature
scFv	Single-chain variable fragment
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
sec	Second
SH2	Src-like homology 2
SL	Sublibrary
ssDNA	Single-stranded deoxyribonucleic acid
STAT	Signal transducer and activator of transcription
TB	Terrific Broth
TMB	3,3',5,5'-tetramethylbenzidine
TNF-α	Tumor necrosis factor-α
TRIM	Trinucleotide-directed mutagenesis
trP1	Truncated P1
VDJ segments	Variable (V), diversity (D) and joining (J) germline gene segments
VEGFR-1	Vascular endothelial growth factor receptor 1
V_H	Variable heavy domain
V_L	Variable light domain
WT	Wild type

1. INTRODUCTION

The biologics market encompasses a large variety of products including recombinant proteins, vaccines, blood products, gene and cellular therapies, etc. Monoclonal antibodies represent the most important class of protein therapeutics in the biologics market (Frenzel *et al*., 2016). Global sales revenue of antibody-based products such as monoclonal antibodies, antibody fragments, Fc-fusion proteins and antibody-drug conjugates was estimated to be $75 billion in 2013, accounting for ~50% of the total sales of all biologics. The global market for antibody-based products is valued at $94 billion in 2017 and is expected to reach nearly $125 billion by 2020 (Ecker *et al*., 2015).

The clinical development of monoclonal antibodies started in the early 1980s, and the first therapeutic monoclonal antibody, Muromonab-OKT3, was approved in 1986 (Emmons and Hunsicker, 1987; Ecker *et al*., 2015). As of March 2017, 63 antibody-based products have been approved by the Food and Drug Administration (FDA) or European Medicines Agency (EMA) (www.antibodysociety.org). Six therapeutic monoclonal antibodies were granted first marketing approvals in 2016 by FDA or EMA. As of December 2016, over 230 antibody-based products are in phase II clinical studies, 52 antibody-based products are in phase III clinical studies for cancer (20 products) or non-cancer indications (32 products), and 9 antibody-based products (8 monoclonal antibodies and 1 antibody-drug conjugate) are under regulatory review by FDA or EMA (Reichert, 2017).

Antibody-based molecules have evolved to become a highly-established class of therapeutics, fueled by progress in target discovery and validation, advances in antibody isolation, engineering and production, and successes in clinical development and commercialization (Scott *et al*., 2012; Freise and Wu, 2015). Also, antibody-based molecules possess favorable biochemical, biophysical, biological, pharmacological and clinical properties for drug development, and have a higher probability of success in acquiring regulatory approval than other drug formats (Aggarwal, 2009; Nelson *et al*., 2010; Chan and Carter, 2010; Scott *et al*., 2012). While the hybridoma technology laid the foundation for modern day antibody therapeutics, the emergence of phage and other display technologies revolutionized the discovery and optimization of antibody-based molecules. A detailed knowledge on structural, functional and pharmacological properties of antibodies allowed researchers to rationally re-engineer lead antibodies for improved stability, potency and safety. The adoption of efficient platform-based

Table 1.1: Therapeutic antibodies approved in 2016/2017 or in review by EMA or FDA

Name	Target	Format	Indication	EU Status	US Status
Atezolizumab	PD-L1	Humanized IgG1	Bladder cancer	Review	2016
Olaratumab	PDGFRα	Human IgG1	Soft tissue sarcoma	2016	2016
Reslizumab	IL-5	Humanized IgG4	Asthma	2016	2016
Ixekizumab	IL-17a	Humanized IgG4	Psoriasis	2016	2016
Bezlotoxumab	*C. difficile* enterotoxin B	Human IgG1	Prevention of *C. difficile* infection	2017	2016
Obiltoxaximab	*B. anthracis* exotoxin	Chimeric IgG1	Prevention of inhalational anthrax	NA	2016
Ocrelizumab	CD20	Humanized IgG1	Multiple sclerosis	Review	2017
Avelumab	PD-L1	Human IgG1	Merkel cell carcinoma	Review	2017
Dupilumab	IL-4Rα	Human IgG4	Atopic dermatitis	Review	2017
Brodalumab	IL-17R	Human IgG2	Psoriasis	Review	2017
Xilonix	IL-1α	Human IgG1	Advanced colorectal cancer	Review	NA
Inotuzumab ozogamicin	CD22	Humanized IgG4 ADC	Hematological malignancy	Review	NA
Sirukumab	IL-6	Human IgG1	Rheumatoid arthritis	Review	Review
Sarilumab	IL-6R	Human IgG1	Rheumatoid arthritis	Review	Review
Guselkumab	IL-23	Human IgG1	Plaque psoriasis	Review	Review
Romosozumab	Sclerostin	Humanized IgG2	Osteoporosis (post-menopause)	NA	Review
Durvalumab	PD-L1	Human IgG1	Bladder cancer	NA	Review
Burosumab	FGF23	Human IgG1	X-linked hypophosphatemia	Review	NA
Benralizumab	IL-5Rα	Humanized IgG1	Asthma	Review	NA
Tildrakizumab	IL-23	Humanized IgG1	Plaque psoriasis	Review	NA
Caplacizumab	Von-Willebrand factor	Humanized nanobody	Acquired thrombotic thrombocytopenic purpura	Review	NA

Table compiled from Reichert, 2017; and www.antibodysociety.org

approaches in antibody bioprocessing and generation of highly productive cell lines for antibody expression enabled large-scale production of antibodies, increased purification yields and decreased cost of goods (Nelson *et al*., 2010; Dübel and Reichert, 2014).

The continued growth and interest in the development of antibody-based products is also fueled by the modularity and versatility of antibodies. Their modular nature resulted in small-sized antibody fragments for developing diagnostic-imaging products, bi-specific antibodies for developing T-cell redirecting therapeutics, and artificial immune-receptors for developing chimeric antigen receptor (CAR) T-cell therapies. The ability to attach various cytotoxic moieties to the antibody framework led to the development of next-generation antibody therapeutics such as antibody-drug conjugates (ADC), antibody-radionuclide conjugates and antibody-cytokine fusions (Chames *et al*., 2009; Weiner, 2015).

The successful development of therapeutic monoclonal antibodies starts with the generation of lead antibodies using hybridoma, transgenic mice or display technologies. Next, lead antibodies are optimized in terms of affinity, specificity, action mechanisms, immunogenicity, effector function, and serum half-life. Antibodies exhibiting desired biochemical and biological properties are then assessed and improved in terms of stability, homogeneity, formulation interactions, pharmacokinetics, safety and efficacy. Antibodies exhibiting favorable production and pharmacological characteristics successfully enter advanced preclinical studies (Shih, 2012; Chiu and Gilliland, 2016). In this PhD project, we describe novel strategies for obtaining lead antibodies from synthetic antibody libraries using the phage display technology. We also describe the generation, isolation, optimization and characterization of synthetic antibody leads against Notch, Jagged and EGF receptors.

2. LITERATURE REVIEW

2.1 Introduction to Antibodies

2.1.1 Components of the Human Immune System

The immune system of vertebrates has evolved a complex network of proteins, specialized cells, tissues and organs to protect us from pathogens. The two main functions of the immune system are antigen recognition and eradication of the source of antigen. The immune system is classified into two main areas: the innate immune system and the adaptive immune system. The innate immune system acts as a first line of defense against pathogens and allows a rapid response to invasion. It has three main components: (1) anatomical barriers including physical (skin and hair), chemical (body fluids) and biological (probiotics) barriers; (2) cell-mediated immunity conferred by leukocytes; and (3) humoral immunity conferred by complement proteins. Leukocytes consist of many cell types including macrophages, neutrophils, eosinophils, basophils, dendritic cells, natural killer cells, mast cells and platelets. Each cell type is specialized to respond, recognize, kill, recruit and activate a certain set of cells. The complement system is a group of about 20 serum proteins that can fulfill a range of immune effector functions (Male *et al*., 2006; Delves *et al*., 2017).

Often, innate immunity is not sufficient for clearing pathogens, and the adaptive immune system is brought into action. In contrast to the innate immune system that senses common molecular patterns in pathogens, the adaptive immune system employs a very large repertoire of genetically programmable recognition proteins for sensing antigens. The use of tunable antigen receptors provides three key advantages: flexibility, specificity and immunological memory. The adaptive immune response is mediated by lymphocytes, which includes T cells and B cells. T cells develop in the thymus and comprise the cell-mediated arm of the adaptive immune system. There are three types of T cells: cytolytic T cells, helper T cells and regulatory T cells. T cell responses require processing and proper presentation of the antigen by major histocompatibility complexes (MHCs) on the surface of antigen-presenting cells (APCs). B cells develop in the bone marrow and comprise the humoral arm of the adaptive immune system. B cells recognize and bind antigens present in serum, tissue fluids or on cell membranes. Upon antigen recognition and activation, B cells differentiate into plasma cells, which secrete large amounts of soluble glycoproteins known as antibodies. Cytolytic T cells and antibodies use different mechanisms to

eliminate pathogens. The innate and adaptive arms of the immune system interact and correlate with each other using APCs (macrophages, dendritic cells and B cells) (Male *et al*., 2006; Delves *et al*., 2017).

2.1.2 Structure and Function of Antibodies

Antibodies, also known as immunoglobulins are glycoproteins that belong to the immunoglobulin superfamily. They constitute most of the gamma globulin component of blood proteins. They are expressed as cell-surface receptors on B cells or as soluble proteins present in serum and tissue fluids. The basic structural unit of an immunoglobulin consists of two identical heavy chains (~50 kDa each) and two identical light chains (~25 kDa each), which are held together by disulfide bridges and non-covalent interactions. There are five different heavy chains (γ, α, μ, δ, ϵ) and two different light chains (κ, λ). The type of the heavy chain determines the class of the antibody. The five distinct classes of immunoglobulin molecules are IgG, IgA, IgM, IgD and IgE. The IgG class is subdivided into four subclasses (IgG1, IgG2, IgG3 and IgG4), and the IgA class is subdivided into two subclasses (IgA1 and IgA2). In total, nine different antibody isotypes are present in all normal individuals (Male *et al*., 2006; Delves *et al*., 2017).

IgG accounts for 70-75% of the total serum immunoglobulin pool, and provides the majority of antibody-based immunity against invading pathogens. It is the only antibody class capable of crossing the placenta and give passive immunity to the fetus. IgA accounts for 15-20% of the total serum immunoglobulin pool. It is predominantly found in mucosal areas, saliva, tears and breast milk, and prevents colonization by pathogens. IgM accounts for 10% of the total serum immunoglobulin pool, and helps to eliminate pathogens in the early stages of antibody-based immunity (before there is sufficient IgG secretion). IgM is also expressed as cell-surface receptors on B-cells. IgD accounts for less than 1% of the total serum immunoglobulin pool, and functions mainly as a cell-surface antigen receptor on naïve B cells. IgE levels are very low in the serum (< 0.05 μg/mL), and it protects against allergens by activating basophils and mast cells (Male *et al*., 2006; Delves *et al*., 2017).

The overall structure of an antibody depends on its class and subclass; nevertheless, there are many common structural features among antibodies. Both light and heavy chains are folded into discrete immunoglobulin domains (~100 amino acids each). These domains have a characteristic immunoglobulin fold in which two beta sheets create a sandwich shape, which are held together by conserved disulfide bridges and non-covalent interactions. The N-terminal

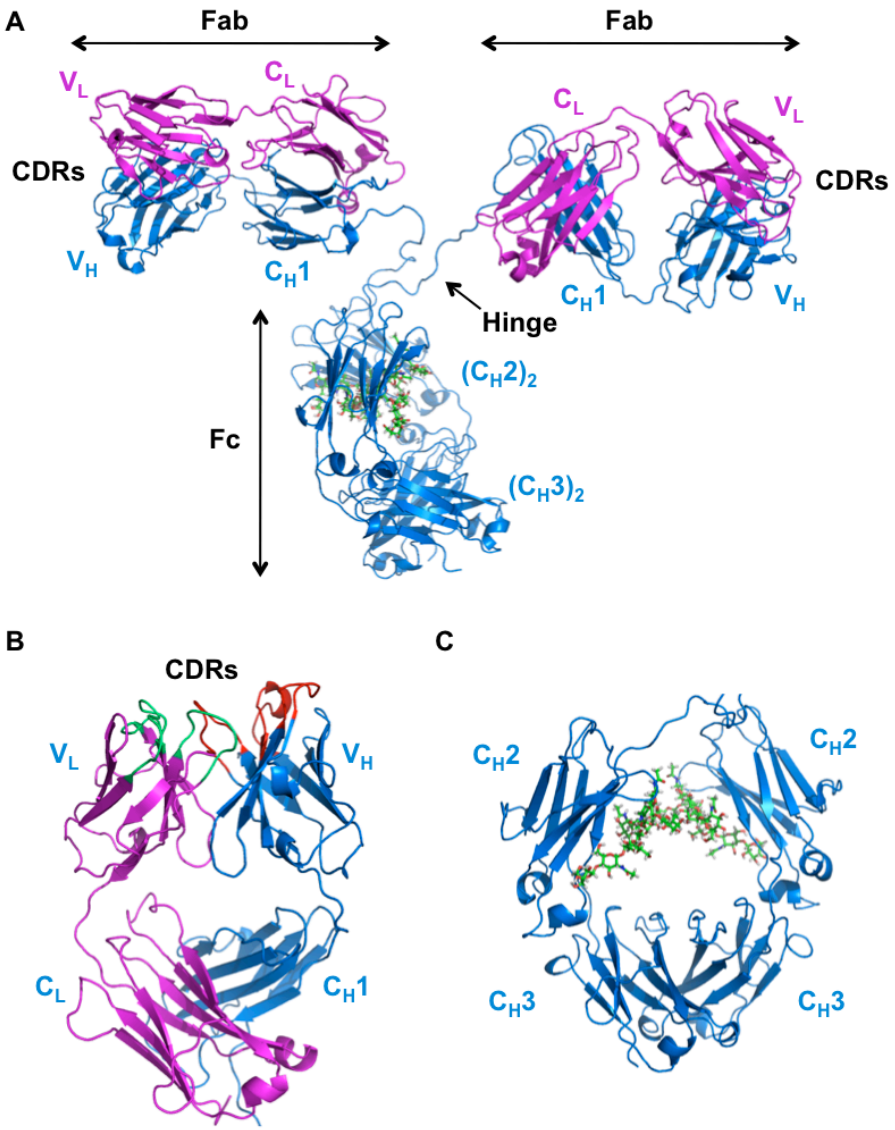


Figure 2.1: Crystal structure of Immunoglobulin G (IgG). (A) IgG molecule is a hetero-tetramer of two light chains (magenta) and two heavy chains (blue). The light chain consists of a variable light domain (V_L) and a constant light domain (C_L) whereas the heavy chain consists of a variable heavy domain (V_H) and three constant heavy domains (C_H1, C_H2 and C_H3). The IgG molecule is held together by multiple disulfide bonds and non-covalent interactions. **(B)** The antigen-binding site is contained within the non-glycosylated antigen-binding fragment (Fab) heterodimer. The Fab includes the variable fragment (Fv), which consists of the variable domains of the light (V_L) and heavy chain (V_H) and can be produced as a single-chain variable fragment (scFv). Antigen recognition is primarily mediated by six hyper-variable loops or complementarity-determining regions (CDRs). Three light chain CDRs (L1, L2 and L3) are colored in green and three heavy chain CDRs (H1, H2 and H3) are colored in red. **(C)** The Fc portion is homo-dimeric and is glycosylated at the conserved Asn297 position located in C_H2. The Fc segment is responsible for binding to Fc receptors (FcγR), neo-natal Fc receptors (FcRn), Complement protein (C1q) and Protein A/G. Cartoon representations of IgG, Fab and Fc molecules are derived from PDB entry 1IGT.

domains of light and heavy chains contain diverse stretches of amino acids known as hypervariable regions, therefore designated as variable domains. Sequences of other light and heavy chain domains are conserved within their own classes, therefore designated as constant domains. The light chain of antibodies contains two domains: one variable light domain (V_L) followed by one constant light domain (C_L). The heavy chain of IgG, IgA, IgD contains four successive domains: one variable heavy domain (V_H) and three constant heavy domains (C_H1, C_H2 and C_H3). The heavy chain of IgM and IgE contains an additional constant domain (C_H4) at its C-terminus. The basic functional unit of each antibody is an immunoglobulin monomer (one Ig unit) composed of two light and two heavy chains. Secreted antibodies can also be dimeric with two Ig units as with IgA or pentameric with five Ig units as with IgM (Male *et al*., 2006; Delves *et al*., 2017).

In the case of IgG, the molecule is 'Y' shaped, and both arms of the Y are flexible due to the presence of an unstructured hinge region between C_H1 and C_H2. Proteolysis of IgG at the hinge region gives rise to three functional fragments: two antigen-binding fragments (Fabs) and one crystallizable fragment (Fc). Each Fab is a heterodimer of a light chain (V_L-C_L) associated with the V_H-C_H1 domains of a heavy chain. The Fc segment is a homodimer of two C_H2-C_H3 domains, and is glycosylated at the conserved Asn297 position located in C_H2. The hypervariable regions, also known as complementarity-determining regions (CDRs), form the antigen-binding site in the Fab fragment. The V_L domain contains three CDRs (CDRL1, CDRL2 and CDRL3) and the V_H domain contains three CDRs (CDRH1, CDRH2 and CDRH3). The CDRs and selected framework residues on variable domains are responsible for interactions with antigens and dictate essential properties such as binding affinity and target specificity. The Fc region is responsible for connecting antigen-antibody binding to antibody effector functions. Two important antibody effector functions are antibody-mediated cellular cytotoxicity (ADCC) and complement-dependent cytotoxicity (CDC). ADCC is the lysis of antibody-coated target cells by immune effector cells such as NK cells, macrophages and neutrophils that express cell-surface Fc receptors. ADCC is initiated by the binding of the Fc region of antibodies on target cells to Fc receptors on immune effector cells. In CDC, the Fc region interacts with the serum C1q protein and initiates the complement system, which eventually leads to the lysis of antibody-coated target cells. The Fc-C1q interaction also labels the antibody-coated target cells for destruction by phagocytosis, a process known as opsonization (Male *et al*., 2006; Delves *et al*., 2017).

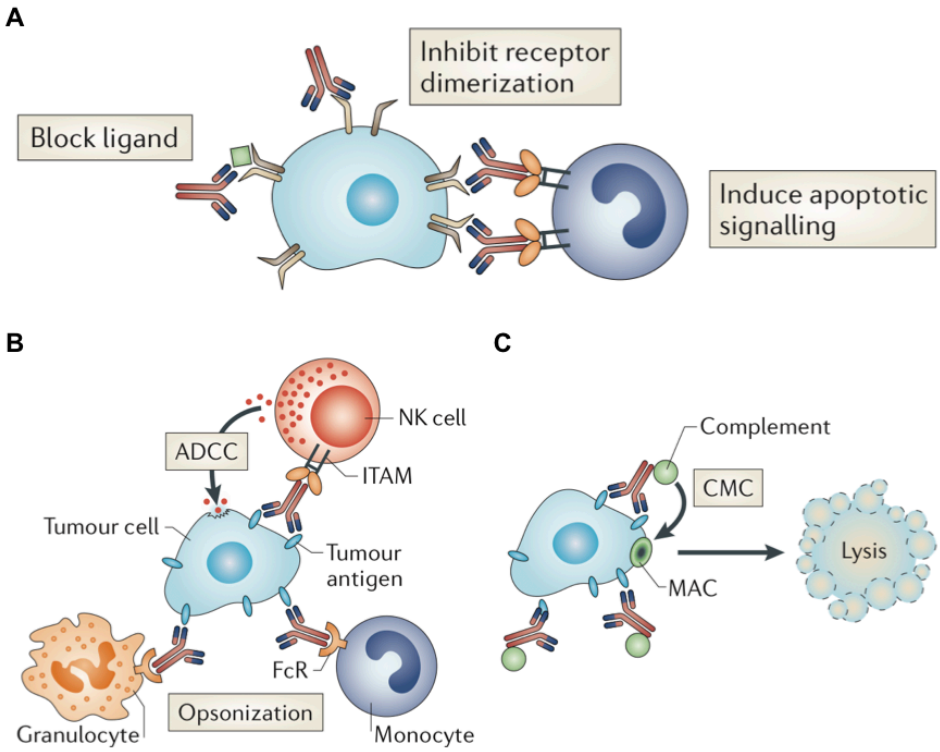


Figure 2.2: General mechanisms of action of therapeutic antibodies. (A) The Fab portion of antibodies can inhibit receptor activation by blocking the binding of an activating ligand to the target receptor and/or by inhibiting the dimerization of cell-surface receptors. Inhibition of receptor activation can cross-link receptors and induce apoptotic signaling. **(B)** The Fc portion of antibodies can mediate antibody-dependent cytotoxicity (ADCC) by binding to immune effector cells such as natural killer cells (NK cells), monocytes, granulocytes and neutrophils, which express Fc receptors (FcγR) and immune-receptor tyrosine-based activation motifs (ITAM). **(C)** The Fc portion of antibodies can mediate complement-dependent cytotoxicity (CDC) by binding to the complement-cascade activating protein C1q. Complement activation results in cell death through development of the membrane attack complex (MAC). Complement fixation also results in opsonization of target cells, which enhances phagocytosis by monocytes and granulocytes. Figure 2.2 reprinted with permission from Nature Publishing Group (Weiner, 2015).

2.1.3 The B-Cell Response

The variable domains of antibodies are formed by somatic recombination of a finite set of tandemly arranged variable (V), diversity (D) and joining (J) germline gene segments. This process, known as VDJ recombination, also results in the addition and deletion of nucleotides at the junctions between ligated germline segments. The V_L domain is encoded by V_L and J_L genes, and the V_H domain is encoded by V_H, D_H, and J_H genes. The first two light and heavy chain CDRs are encoded by the V segment, while the third CDR is the product of V-J regions for CDRL3 and V-D-J regions for CDRH3. Following VDJ recombination, a fully assembled

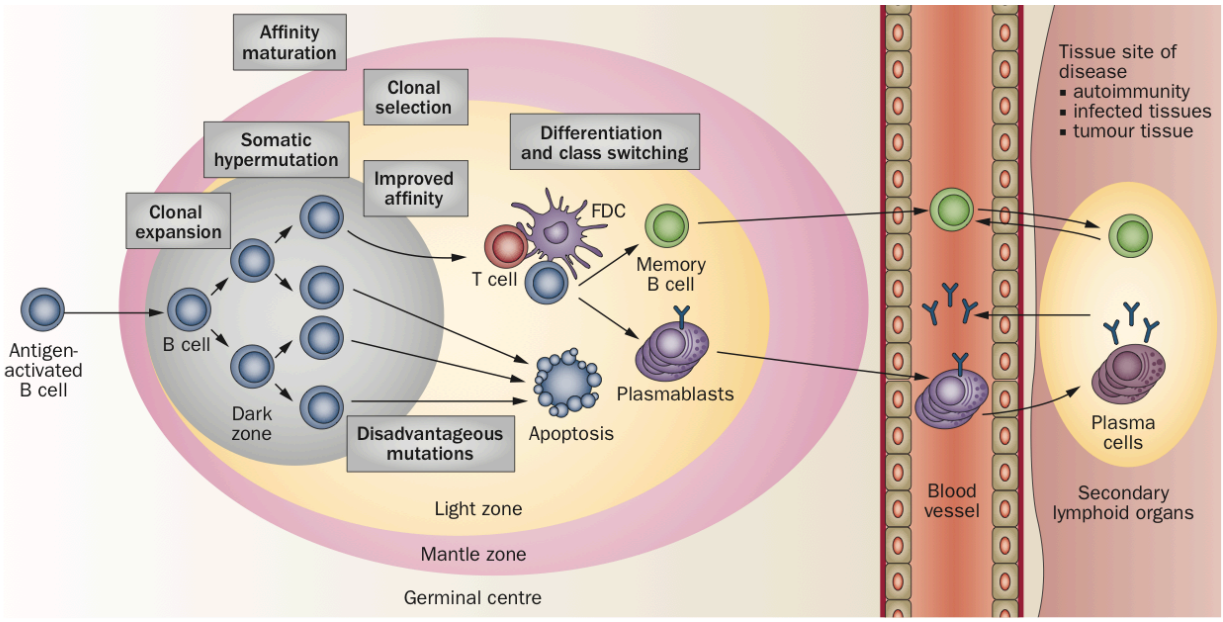


Figure 2.3: The B-cell response. Normal B cells are generated in the bone marrow and exist as IgD+ and IgM+ naïve B cells in peripheral blood. Naïve B cells undergo germinal center reaction after encountering a cognate antigen and receiving assistance from T-cells. The germinal center reaction occurs within lymph nodes and gives rise to multiple clonal families of B cells. The reaction consists of multiple steps: clonal expansion, affinity maturation, class switching and differentiation. Affinity maturation involves two processes: somatic hypermutation and clonal selection. B cells that survive the germinal center reaction ultimately differentiate into plasmablasts, plasma cells and memory B cells. Plasmablasts circulate in the blood, migrate to diseased-tissues and secondary lymphoid organs, and secrete antibodies. Plasma cells reside in the bone marrow and lamina propria, and secrete antibodies. Memory B cells circulate in the blood and differentiate directly into antibody-producing cells upon re-exposure to their cognate antigen. Figure 2.3 reprinted with permission from Nature Publishing Group (Robinson, 2015).

antibody heterodimer (IgD and/or IgM) is expressed on the surface of naïve B cells. When a naïve B cell encounters an antigen in the right environment (T-cell help and co-stimulatory signals), it goes through multiple stages of development to become an antigen-specific B cell that produce antibodies against the antigen. The overall process known as the germinal center reaction consists of the following steps: clonal expansion, affinity maturation, class switching and differentiation. Stimulation of cell-surface IgD and/or IgM molecules by the antigen leads to rapid proliferation of naïve B cells, a process known as clonal expansion. Affinity maturation involves two processes: somatic hypermutation and clonal selection. During somatic hypermutation, antibody variable regions, primarily CDRs, are mutated 1-2 times per cell division. During clonal selection, B cells compete for antigens and growth factors in the germinal

center, and B cells expressing high-affinity antibodies are only selected for further expansion and survival. Following multiple rounds of affinity maturation, B cells bearing high-affinity antibodies for the cognate antigen undergo class-switch recombination to express IgG, IgA or IgE isotypes. Class-switched B cells further respond to growth factors and other signals and differentiate into antibody-producing plasmablasts, plasma cells and memory B cells. Plasmablasts migrate to tissues involved in the disease process and secondary lymphoid organs, and secrete antibodies. Plasma cells reside primarily in the bone marrow and secrete antibodies. Memory B cells circulate in the blood and differentiate directly into antibody-producing cells upon re-exposure to their target antigen (reviewed by Victora and Nussenzweig, 2012; Gitlin *et al*., 2014; Georgiou *et al*., 2014; Robinson, 2015).

2.2 Generation of Antibodies

2.2.1 Hybridoma and Display Technologies

Antibodies can be generated using hybridoma-based technologies and display technologies. Kohler and Milstein first described the classical hybridoma approach in 1975 and received a Nobel Prize for their discovery in 1984. This technology requires the fusion of immortal myeloma cells with antibody-producing B-lymphocytes from an immunized animal (Kohler and Milstein, 1975). The resulting hybrid cell acquires the capacity of indefinite growth from the myeloma tumor cell and the capacity to produce a specific antibody from the B-lymphocyte cell. The hybridoma cell secreting the desired antibody is isolated from the population of fused hybrid cells, stabilized by repeated cell cloning, and used to produce fairly large quantities of identical antibody for years (Moldenhauer, 2014). The hybridoma technology revolutionized both basic and applied sciences by providing monoclonal antibodies that served as a key tool for researchers, as tracking and detection reagents in medical laboratories, and also as therapeutics in the hands of clinicians (Nelson *et al*., 2010). From a clinical standpoint, hybridoma antibodies are typically derived from murine sources, which bear the potential for triggering hypersensitivity reactions and anti-mouse antibodies in humans (Pendley *et al*., 2003; Schmidt *et al*., 2009). Strategies such as chimerization and humanization have been devised to reduce the immunogenicity of monoclonal antibodies (Almagro and Fransson, 2008), and transgenic mouse strains have been developed to produce human antibodies instead of mouse antibodies (Lonberg, 2008). However, bottlenecks in the immunization process, intense manual

liquid handling during generation and characterization of antibodies, and the lack of direct access to antibody genes have slowed down antibody development. Further, it is difficult to generate antibodies against unstable, toxic and highly conserved targets in animal systems (Sidhu, 2012).

Fortunately, the advent of *in vitro* display technologies has enabled generation, identification and engineering of fully-human antibodies without the use of animals (Geyer *et al*., 2012). Antibody display technologies can be roughly divided into cell-free (phage and ribosome display) and cell-based (yeast, bacterial and mammalian cell display) display technologies (Harel-Inbar and Benhar, 2012). Though several display systems are available, the basic principle behind these methods centers on the coupling of genotype (antibody genes) with phenotype (binding to proteins). The availability of numerous display platforms offers different ways to generate, manipulate and present antibodies to targets during binding selections (Geyer *et al*., 2012). Phage display is the most robust and well-established of these methods and has yielded numerous antibodies for research and clinical applications (Shim, 2016). In antibody engineering, phage display is primarily used to isolate antigen-specific antibody fragments from Fab or scFv libraries. In contrast to animal immunization, phage display offers precise control over selection conditions, for example, presentation of specific conformations of the target or addition of competitors to direct selections against epitopes of interest (Sidhu and Fellouse, 2006; Bradbury *et al*., 2011). Further, the gene encoding the antibody is cloned simultaneously with selection, which provides many advantages for engineering and rapid characterization of selected antibodies such as improving the affinity and stability of antibodies, altering the specificity of antibodies and sub-cloning antibody genes into alternate expression platforms (Michnick and Sidhu, 2008; Geyer *et al*., 2012). Also, phage display is wholly amenable to integration with robotic liquid-handling devices and highly sophisticated equipment for antibody selection, production and characterization, which are essential for high-throughput generation of antibodies (Miersch and Sidhu, 2012).

2.2.2 Phage Display of Antibody Fragments

Although many viruses (including λ, T4 and T7) have been employed for phage display, most antibody phage display platforms use M13- *E. coli* phage-host system because of the following reasons: (1) the non-lytic nature of phage infection, (2) high viral titer capacity, (3) simultaneous presence of both single- and double-stranded forms of viral DNA, (4) little size constraint on inserted DNA and (5) assembly of display proteins in the bacterial periplasm

suitable for disulfide bond formation. M13 belongs to the Ff class of filamentous bacteriophages (genus *Inovirus*) that infects a wide variety of gram-negative bacteria. Ff viral particles are pencil-shaped; ~900 nm long and ~65 Å in diameter. The single-stranded genome is made up of ~6400 bases that encodes a total of 11 phage proteins. Six proteins (pI, pII, pIV, pV, pX and pXI) are involved in phage DNA replication, phage assembly and secretion, and five proteins (pIII, pVI, pVII, pVIII and pIX) form the capsid that encloses the single-stranded phage DNA. The phage filament is comprised of ~2700 copies of the major coat protein pVIII, which is capped on both ends by ~5 copies of minor coat proteins (pIII and pVI on one end, and pVII and pIX on the other end). Filamentous phage infection is a multi-step process; recognition of the bacterial cell-surface receptor F-conjugative pilus by the phage pIII protein, followed by pilus retraction, binding of the pIII protein to the periplasmic TolA receptor, uncoating of viral DNA and its concomitant translocation into the host cell cytoplasm. Inside the host cell, single-stranded DNA (ssDNA) is converted to a double-stranded replicative form by host polymerases, which then undergoes translation to enable synthesis of new viral proteins and rolling-circle replication to enable synthesis of progeny ssDNA. Nascent viral coat proteins exist as integral membrane proteins in the inner membrane of *E. coli*. One-hour post-infection, the pV protein pre-packages the new ssDNA for viral assembly, packaging and extrusion. Viral assembly and export occurs through a membrane pore without host cell lysis (reviewed by Rodi *et al*., 2005; Miersch and Sidhu, 2012; Rodi *et al*., 2015).

In 1985, George Smith first demonstrated that functional polypeptides could be displayed on the surface of phage particles as fusions to the pIII protein (Smith, 1985). Shortly thereafter, several groups showed that Fab or scFv could also be displayed on phage particles and led the way for the development of phage-displayed antibody libraries (Huse *et al*., 1989; Barbas *et al*., 1991; McCafferty *et al*., 1990; Clackson *et al*., 1991). Since then, phage display is constantly being applied to new experimental and practical problems, giving rise to many phage display reagents, methods and platforms. Here, we will restrict the discussion to the phage display system used is this project. We utilized a phagemid- helper phage system that has been optimized for the efficient construction and screening of Fab libraries displayed on M13 phage particles as fusions to a truncated pIII protein. Since integration of exogenous display proteins into the viral genome hampers phage infectivity, production and stability; historically, phage display platforms have utilized a phagemid- helper phage system (Miersch and Sidhu, 2012). The phagemid is a

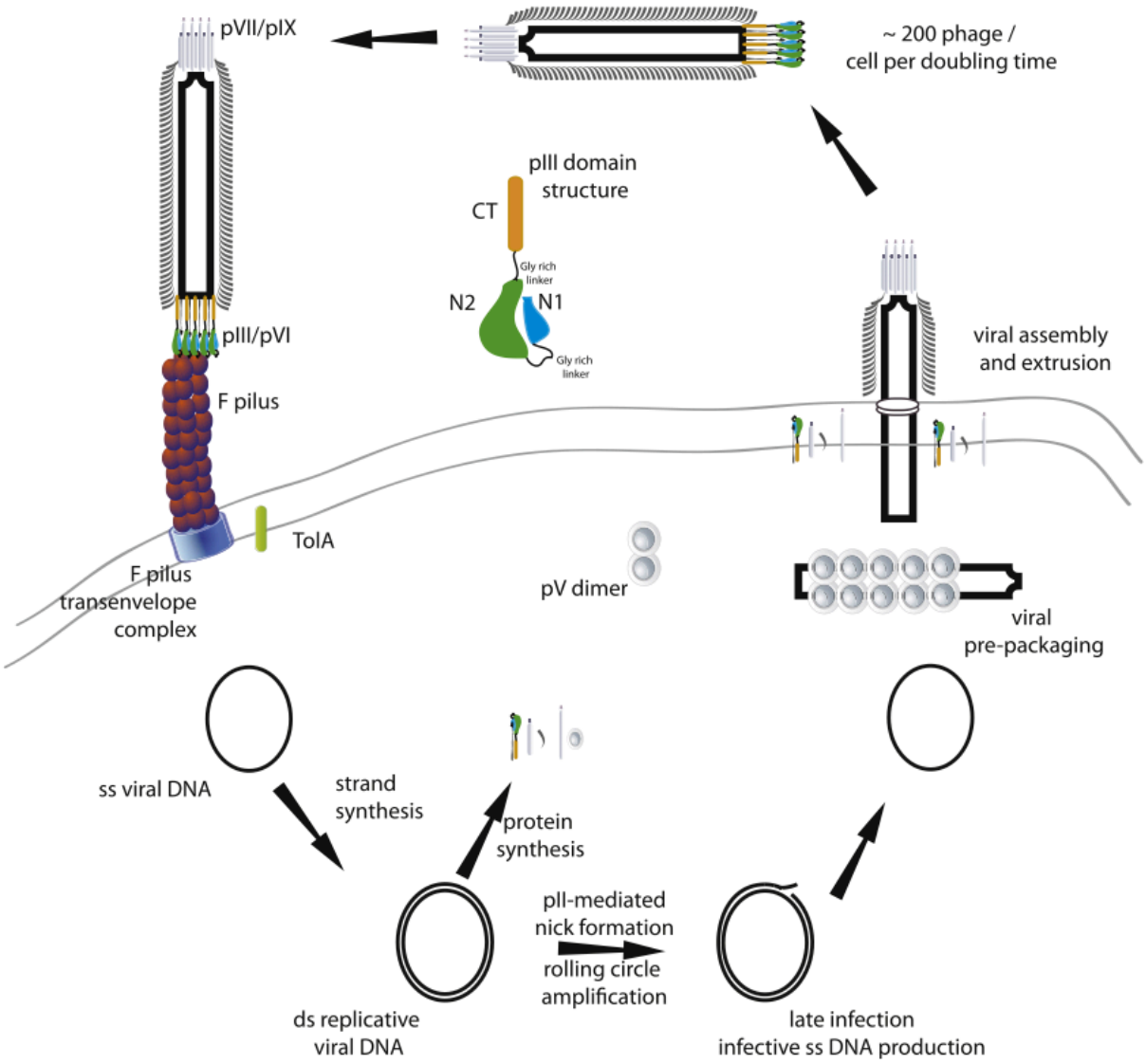


Figure 2.4: M13 phage life cycle. M13 phage particles initiate contact with gram-negative bacteria expressing the F conjugative pilus using the pIII coat protein. The pIII protein that mediates the infection process is composed of three domains: N1 and N2 domains at the N-terminus and the CT domain at the C-terminus. During infection, the N2 domain interacts with the F-pilus on the host bacterium. The F-pilus retracts and enables binding between the N1 domain and the TolA receptor in the bacterial periplasm. Through additional interactions, pIII is inserted into the bacterial inner membrane, and the CT domain mediates the release of phage ssDNA into the bacterial cytoplasm. Phage ssDNA is converted into the dsDNA replicative form by bacterial polymerases. This dsDNA form is used for the synthesis of new phage proteins that translocate and reside in bacterial inner-membrane in preparation for viral assembly. The dsDNA also undergoes rolling circle amplification for the generation of new ssDNA. One-hour post-infection, the pV protein pre-packages the new ssDNA for viral assembly, packaging and extrusion. Viral assembly and export occurs through a membrane pore without host cell lysis. Figure 2.4 reprinted with permission from Elsevier (Miersch and Sidhu, 2012).

double-stranded plasmid designed to express the exogenous display protein- coat protein fusion within the bacterial host. The phagemid used in this work (pHP153) contains a truncated form of *gIII*, an antibiotic resistance marker and double-stranded DNA origin of replication for phagemid propagation in *E. coli,* and an f1 origin of replication for synthesis and packaging of ssDNA (Lee *et al*., 2004A; Fellouse and Sidhu, 2006). The truncated *gIII* only expresses the C-terminal domain of pIII required for anchoring the coat protein into the viral capsid and for the assembly of phage particles. The two N-terminal domains (N1 and N2) required for host infection are not fused to the Fab protein. Since, the phagemid lacks wild-type pIII and other phage genes, functional phage production requires co-infection with helper phage M13KO7 that contains all phage genes (Miersch and Sidhu, 2012; Rajan and Sidhu, 2012). M13KO7 contains a point mutation (Met40Ile) in *gII*, which substantially decreases the replication of M13KO7 ssDNA and preferentially packages the phagemid DNA into phage particles (Russell *et al*., 2004).

We chose to use Fab over other antibody formats because a higher proportion of Fabs in the library are stable, functionally folded, well displayed, easily expressed and purified and reliably converted into functional bivalent IgG molecules (Ponsel *et al*., 2011; Miersch and Sidhu, 2012; Shim, 2016). The pHP153 phagemid is a bicistronic vector that expresses light and heavy chains of Fabs as two open-reading frames under the control of an alkaline phosphatase promoter. Truncated *gIII* is fused to the C-terminus of the Fab heavy chain. The phagemid also contains leader sequences at the N-terminus of both chains to export the Fab light chain and the Fab heavy chain-truncated pIII fusion via the Secretory pathway. Disulfide bond-mediated folding of light and heavy chains occur in the oxidizing environment of the bacterial periplasm. Both chains associate non-covalently to form Fab molecules (Fellouse and Sidhu, 2006; Rajan and Sidhu, 2012). Following helper phage infection, synthesized phage coat proteins reside in the inner membrane of *E. coli* as integral membrane proteins. The pV protein pre-packages the phagemid ssDNA for phage assembly, packaging and extrusion. During phage assembly, if the Fab-truncated pIII fusion protein is incorporated into the viral capsid along with other coat proteins including wild-type pIII, the resulting phage particle displays Fabs on its surface in addition to encapsulating its phagemid ssDNA. If all five copies of pIII in the viral capsid contain the wild-type protein, the resulting phage particle does not display any Fabs. Around 200-300 phage particles are produced per *E. coli* per doubling time (~1000 per hour) (Miersch and Sidhu, 2012; Shim, 2016). The phagemid also supports bivalent display of Fabs to resemble

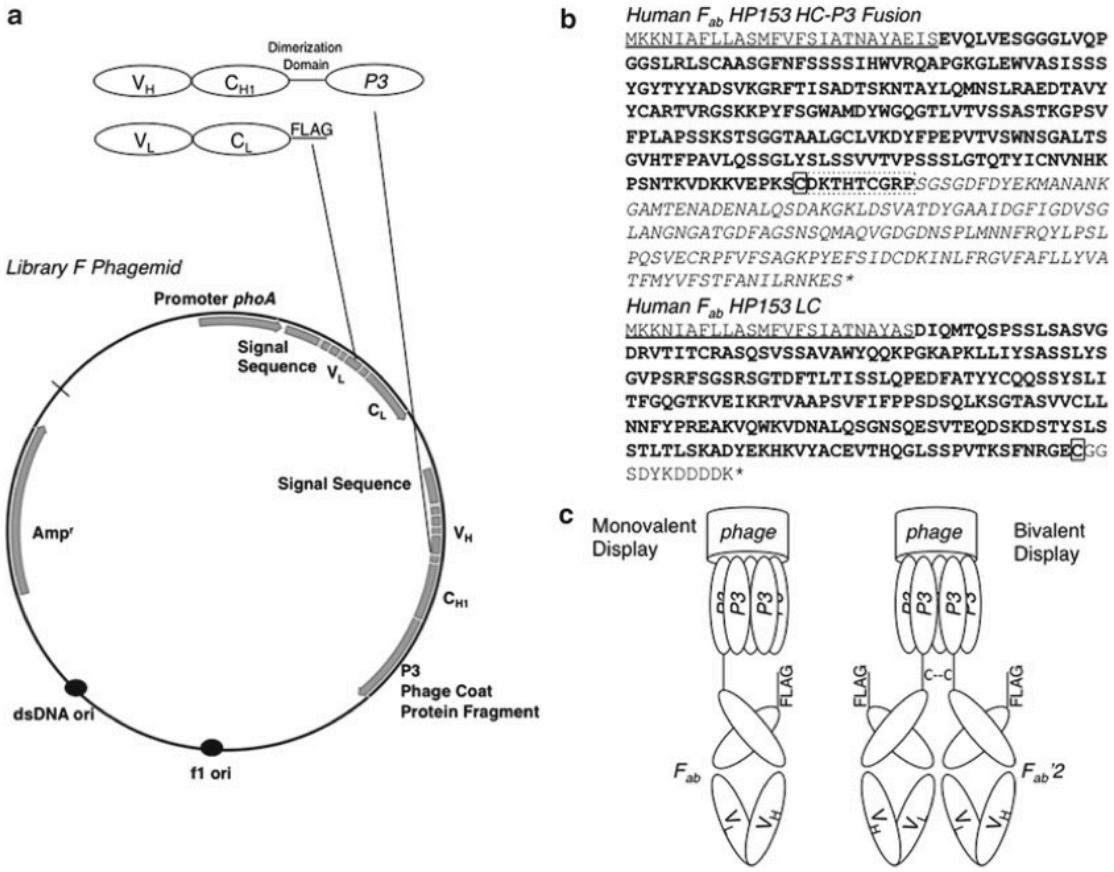


Figure 2.5: Phage display of Fabs in the bivalent format. (A) Schematic of the pHP153 phagemid used for bivalent Fab display. A phosphatase-A (phoA) promoter drives the bicistronic expression of light and heavy chain fragments. The light chain contains the V_L domain, the C_L domain and a C-terminus FLAG tag. The heavy chain contains the V_H domain, the C_H1 domain, a dimerization domain and the truncated pIII protein. The N-terminus secretion signals direct the light and heavy chains to the bacterial periplasm where they associate to form Fabs. The phagemid contains origins of replication for single-stranded DNA (f1 ori) and double-stranded DNA (dsDNA ori), and a selection marker (Amp^r) that confers resistance to carbenicillin. **(B)** Amino acid sequence of the anti-maltose-binding protein Fab heavy and light chain fragments encoded by the template phagemid. Signal sequences are underlined, Fab encoding sequences are in bold letters and the truncated PIII protein is in italics. Cysteines that form the inter-chain disulfide bond between the heavy and light chains are shown in solid boxes. The dimerization domain is shown in a dashed box. The cysteine within the dimerization domain is utilized for displaying Fabs in the bivalent format. (**C**) Schematic of monovalent (Fab) and bivalent (Fab^2) display arrangements of Fabs on M13 phage particles. The disulfide bond between two heavy chain dimerization domains is shown for the bivalent display format. Figure 2.5 reprinted with permission from Springer (Adams *et al*., 2014).

the antigen-binding site of an IgG antibody. Bivalent Fab display is achieved by engineering a cysteine-containing peptide at the linker region that fuses the C-terminus of Fab CH1 with truncated pIII. Disulfide bond formation between two Fab heavy chain-truncated pIII fusion proteins joins two Fab molecules at the hinge region (Lee *et al*., 2004A; Adams *et al*., 2014).

2.2.3 Phage-Displayed Antibody Libraries

Recombinant antibody libraries used in phage display are classified into natural, synthetic and semi-synthetic libraries. The fundamental difference between these libraries lies in their source of diversity. Natural libraries are derived from donor-derived B-lymphocytes (biological diversity), whereas synthetic libraries are assembled from synthetic genes and oligonucleotides (chemical diversity). Semi-synthetic libraries incorporate diversity by combining material derived from both sources (Ponsel *et al*., 2011; Shim, 2016).

Natural repertoires are classified into naïve and immune antibody libraries. They are generated using B-cells extracted from unimmunized animals, and animals immunized with an antigen of interest, respectively. To construct natural antibody libraries, antibody variable regions are amplified from B-cell cDNA by PCR and assembled into a phagemid vector (Hoogenboom, 2005). Naïve libraries usually contain a large diversity (up to 10^{11} members) and can be used to generate antibodies against a wide variety of antigens. Since immune libraries are generated from immunized donors, library members are predisposed for recognition of certain antigens, thus not suitable for the identification of antibodies against a large panel of antigens. Immune libraries therefore contain lower sequence diversity (10^7 members) than naïve libraries (Ponsel *et al*., 2011). Naïve antibody libraries have been used successfully to generate therapeutic antibodies such as Adalimumab, Ramucirumab, Belimumab, Raxibacumab and Necitumumab (Nixon *et al*., 2014; Frenzel *et al*., 2016). There are also a few limitations with using natural antibody repertoires. Since light and heavy chain genes are amplified from different chromosomal locations, their combinatorial assembly introduces random pairing of heavy and light chain variable regions. This results in the reduction of functional antibody diversity in the library, and leads to the isolation of antibodies with unusual V_H-V_L pairs and sub-optimal biophysical properties (Benhar, 2007; Finlay and Almagro, 2012). It is difficult to obtain antibodies against self-antigens from naïve libraries because such antibodies are often deleted by the immune system to prevent autoimmune disorders. The need for immunization restricts the

Table 2.1: Phage display-derived monoclonal antibodies approved for therapeutic applications or in phase III clinical studies

Name	Target	Phage library, Format	Final Format	Indication	Status
Adalimumab	TNF-α	Semi-synthetic, scFv	IgG1	Rheumatoid arthritis Crohn's disease Psoriatic arthritis Plaque psoriasis	Approved 2002
Belimumab	B-LyS	Naïve, scFv	IgG1	Systemic lupus erythematosus	Approved 2011
Raxibacumab	*B. anthrasis* Antigen	Naïve, scFv	IgG1	Prophylaxis Anthrax prevention	Approved 2012
Necitumumab	EGFR	Naïve, Fab	IgG1	Non-small cell lung-cancer	Approved 2015
Avelumab	PD-L1	Naïve, Fab	IgG1	Merkel cell carcinoma	Approved 2017
Bimagrumab	ACVR2B	Synthetic, Fab	IgG1	Cachexia Sporadic inclusion body myocitis	Phase III
Briakinumab	IL-12/23	Naïve, scFv	IgG1	Psoriasis Multiple sclerosis Crohn's disease	Phase III
Darleukin	Fibronectin	Semi-synthetic, scFv	ScFv-IL2	Melanoma	Phase III
Gantenerumab	Amyloid β	Synthetic, Fab	IgG1	Alzheimer's disease	Phase III
Ganitumab	IGF1R	Naïve, Fab	IgG1	Pancreatic cancer	Phase III
Guselkumab	IL-23	Synthetic, Fab	IgG1	Plaque psoriasis	Phase III
Lanadelumab	Kallikrein	Naïve, Fab	IgG1	Hereditary angioedema	Phase III
Tralokinumab	IL-13	Naïve, scFv	IgG4	Asthma Ulcerative colitis	Phase III

Table compiled from Frenzel *et al*., 2016; Shim, 2016; and www.antibodysociety.org

development of fully human immune libraries, and antibodies isolated from animal-derived immune libraries provoke immune responses in patients (Ackerman *et al*., 2011).

Semi-synthetic libraries represent some of the earliest attempts to generate antibodies from libraries constructed using non-natural sources. They are part natural and part synthetic in a variety of formats. For example, semi-synthetic libraries can have natural V_H and synthetic V_L, natural frameworks and synthetic CDRs, and synthetic frameworks and natural CDRs (Hoogenboom, 2005; Shim, 2016). Synthetic libraries are constructed by *in vitro* diversification

of CDRs within chosen antibody light and heavy chain variable domains. They were developed to carefully control the usage of antibody frameworks and the composition of CDR diversity (Shim, 2015). Also, chemical synthesis of antibody frameworks enables the incorporation of desirable features for subsequent display and characterization, and CDR diversities encoded by synthetic oligonucleotides are not limited to the scope of the natural immune system. The use of optimal antibody frameworks and tailored CDR diversity makes synthetic antibody libraries ideal for therapeutic antibody discovery (Harel-Inbar and Benhar, 2012; Adams and Sidhu, 2014).

Synthetic repertoires are classified into multiple- and single-framework synthetic antibody libraries (Ponsel *et al*., 2011). In multiple framework libraries, many antibody frameworks are chosen and used for library construction. For example, the HuCAL (Human combinatorial antibody library) series of synthetic libraries uses up to seven V_H and seven V_L genes resulting in 49 different V_H-V_L combinations that augment the antibody variable domain diversity. Since framework residues are known to influence CDR conformations, the presence of multiple frameworks also enables the engineering of CDRs with a wide variety of canonical conformations (Knappik *et al*., 2000; Rothe *et al*., 2007; Prassler *et al*., 2011; Tiller *et al*., 2013). Multiple framework libraries are advantageous in terms of antibody framework diversity and CDR conformational diversity, however construction of multiple-framework libraries is very expensive, time-consuming and typically undertaken in an industry environment (Shim, 2015). The observation that one or two frameworks are often overrepresented after selection from multiple-framework synthetic antibody libraries, together with the desire to increase the functional diversity of the synthetic antibody library by structure-inspired design led to the development of single-framework synthetic antibody libraries (Ponsel *et al*., 2011). Single-framework libraries are constructed by targeted diversification of selected CDR residues of a single chosen antibody framework (Benhar, 2007). Single-framework libraries offer many advantages over multiple-framework libraries. First, libraries can be built on a single, clinically validated, human antibody framework that has optimal biophysical and pharmacokinetic properties. Second, CDR diversities can be designed based on known structures of the framework, which increases the number of stable and functional members in the antibody library. Third, design and construction of single-framework libraries are relatively easy and less expensive, and they provide valuable information on the fundamentals of antigen-antibody

interactions. Fourth, single framework libraries are well suited for high-throughput antibody generation pipelines, as the defined nature of the framework enables rapid sequence analysis and downstream characterization, and allows facile reformatting between different vector systems for affinity maturation and antibody expression (Miersch and Sidhu, 2012; Adams and Sidhu, 2014; Shim, 2015).

Though synthetic antibody libraries offer many advantages, natural repertoires have been used extensively for phage display selections, as simple exploitation of natural repertoires is considered easier than constructing useful synthetic repertoires. Synthetic antibody libraries are built from scratch, and therefore this technology requires detailed and extensive knowledge of antibody structure and function and antigen recognition by antibodies. In addition, there is a need for highly sophisticated equipment to design and implement the synthetic antibody technology (Fellouse and Sidhu, 2005; Fellouse and Sidhu, 2015). Despite these issues, there has been a significant progress in the field of synthetic antibody engineering. Many synthetic antibody libraries have been constructed and used successfully in antibody discovery programs, and a few synthetic antibodies including Bimagrumab, Gantenerumab, Guselkumab and Avelumab have entered phase III clinical trials (Nixon *et al*., 2014; Frenzel *et al*., 2016).

The Sidhu lab at the University of Toronto has made major contributions towards the development of single-framework synthetic antibody libraries. His group optimized phage vectors and antibody frameworks to ensure libraries are well displayed as functional proteins on phage surfaces (Sidhu *et al*., 2000A; Fuh and Sidhu, 2000; Roth *et al*., 2002; Held and Sidhu, 2004; Sidhu *et al*., 2007; Barthelemy *et al*., 2008). They also developed robust methods for construction, screening and characterization of phage-displayed Fab libraries that contain $>10^{10}$ members (Sidhu *et al*, 2000B; Vajdos et al., 2002; Sidhu and Weiss, 2003; Fellouse and Sidhu, 2006; Pal *et al*., 2006). Phage-displayed libraries constructed by his group have been used to develop antibodies against numerous proteins, providing valuable reagents and potential therapeutics (Sidhu *et al*., 2004; Lee *et al*., 2004B; Fellouse *et al*., 2007; Persson *et al*., 2013; Ma *et al*., 2013). In many cases, they conducted detailed structural and functional studies of engineered antibodies to understand the molecular basis for antigen recognition and they used this knowledge to further improve antibody function (Fellouse *et al*., 2006; Sidhu and Koide, 2007; Birtalan *et al*., 2008; Fisher *et al*., 2010; Reshetnyak *et al*., 2013). In addition to providing reliable methods, libraries and antibodies, these studies resulted in the accumulation of a large

Table 2.2: Representative examples of phage-displayed antibody libraries

Library name	Type, Format	Note	Lib. size	Reference
CAT-BMV	Naïve, scFv	Derived from bone marrow, PBL and tonsils	1.4×10^{10}	Vaughan *et al*., 1996
CAT-DP47	Naïve, scFv	Derived from spleens	1×10^{10}	Groves *et al*., 2006
CAT-CS	Naïve, scFv	Derived from spleen and fetal liver	1.2×10^{11}	Lloyd *et al*., 2009
HAL9/10	Naïve, scFv	Derived from PBL	1.5×10^{10}	Kugler *et al*., 2015
Anti-HIV gp120	Immune, Fab	Derived from HIV-positive patients	1×10^{7}	Burton *et al*., 1991
Anti-acetylcholine receptor	Immune, Fab	Derived from Myasthenia Gravis patients	1.1×10^{6}	Graus *et al*., 1997
Anti-HER2	Immune, Fab	Derived from colorectal cancer patients	2×10^{7}	Clark *et al*., 1997
Anti-Ebola virus	Immune, Fab	Derived from Ebola-positive patients	6×10^{6}	Maruyama *et al*., 1999
n-CoDeR	Semi-synthetic, scFv	Natural CDRs on a synthetic framework	2×10^{9}	Söderlind *et al*., 2000
Dyax	Semi-synthetic, Fab	Synthetic V_H, natural CDRH3 and V_L	3.5×10^{10}	Hoet *et al*., 2005
HuCAL	Synthetic, scFv	Multiple frameworks TRIM CDR	2.1×10^{9}	Knappik *et al*., 2000
HuCAL Gold	Synthetic, Fab	Multiple frameworks TRIM CDR	1.6×10^{10}	Rothe *et al*., 2008
HuCAL Platinum	Synthetic, Fab	Multiple frameworks TRIM CDR	4.5×10^{10}	Prassler *et al*., 2011
Ylanthia	Synthetic, Fab	Multiple frameworks Slonomics CDR	1.3×10^{11}	Tiller *et al*., 2013
Rajpal lab	Synthetic, Fab	Multiple frameworks Slonomics CDR	3.6×10^{10}	Zhai *et al*., 2011
Library-D	Synthetic, Fab	Single framework Tailored and TRIM CDR	3×10^{10}	Fellouse *et al*., 2007
Library-F	Synthetic, Fab	Single framework Tailored and TRIM CDR	3×10^{10}	Persson *et al*., 2013
Shim lab	Synthetic, scFv	Single framework CDR by array synthesis	8×10^{8}	Bai *et al*., 2015

knowledge base relating to antibody structure and function, which can be readily applied to the development of improved synthetic antibody libraries.

2.2.4 Design, Construction and Screening of Synthetic Antibody Libraries

Synthetic antibody libraries differ in terms of size, framework and CDR composition, and method of preparation. The size of an antibody library is commonly described using three terms; theoretical, practical and functional diversity. The theoretical diversity is the number of combinatorial possibilities encoded by the oligonucleotides used for randomization. The practical diversity is the maximum size of the library that can be achieved by the display technology. In phage display, the practical diversity is 10^{10} clones due to limitations in the transformation efficiency of *E. coli* used for library construction. The functional diversity is the number of antibody clones that are folded and displayed properly and are capable of binding. In a typical antibody library, the theoretical diversity is higher than the practical diversity, and the functional diversity is lower than the practical diversity. Synthetic antibody libraries are designed and constructed to display sufficient functional diversity for generating antibodies with favorable biophysical properties against a broad range of antigens (Reviewed by Ponsel *et al*., 2011; Miersch and Sidhu, 2012).

Phage-displayed antibody libraries with 10^{10} members can completely cover the sequence space of only six sites randomized with twenty amino acids. With an optimal length for CDRL3 and CDRH3 loops, the antibody framework that we have chosen contains 63 sites for randomization in its six CDRs. Consequently, it is impossible to generate a phage display library that fully covers the combinatorial diversity of these sites ($\sim 10^{82}$). Therefore, while designing synthetic antibody libraries, one has to limit the number of CDR positions or the number of amino acid types used at a given position without compromising the binding function of antibodies (Sidhu and Kossiakoff, 2007; Miersch and Sidhu, 2012). Analysis of antibody sequences and structures has revealed the CDR positions and amino acid types that have greater effects in antigen binding. The CDRs do not play equal roles in antigen binding, and there is a distinct hierarchy among the CDRs in this regard. Heavy chain CDRs dominate over their light chain counterparts, and among the heavy chain CDRs, CDRH3 plays the major role in determining antibody specificity and affinity (Zemlin *et al*., 2003; Birtalan *et al*., 2008; Burkovitz and Ofran, 2016). Within CDRs, anchor residues are critical for maintaining the canonical conformations of CDR loops, and solvent exposed residues are critical for mediating

interactions with the antigen (Chothia *et al*., 1992). Once the desired CDRs and CDR positions are identified, the sequence diversity for each CDR position has to be precisely defined. Fortunately, in human antibodies >90% of the sequence diversity for any one CDR position can be accounted for in most cases by approximately five to six amino acids (Kabat *et al*., 1991; Ramaraj *et al*., 2012).

The Sidhu lab made a major breakthrough in balancing requirements for CDR residue coverage with sequence diversity. A series of antibody libraries based on "reduced genetic codes", which employ only a fraction of the natural amino acid types, were constructed and screened by phage display to produce high-affinity Fab and scFv molecules (Fellouse *et al*., 2004; Sidhu and Kossiakoff, 2007; Gilbreth *et al*., 2008; Koide and Sidhu, 2009). In the simplest case, a binary code of only two amino acids (tyrosine and serine) has been used to construct minimalist synthetic antibody libraries that have been used to raise specific Fabs against many proteins (Fellouse *et al*., 2005). Recently, a single-framework synthetic Fab library known as library-F was designed and generated in the Sidhu lab, and was used to isolate high-affinity Fabs against different classes of antigens. The library diversity was restricted to solvent-exposed residues within three heavy chain CDRs and CDRL3. Binary amino acid diversity was added to positions within CDRs H1 and H2. Within CDRs L3 and H3, the chemical diversity was generated using nine amino acids at each position, which was biased in favor of tyrosine (25%), serine (20%) and glycine (20%). Further, length diversity was incorporated into the library by allowing different loop lengths for CDRs L3 and H3 (Persson *et al*., 2013; Hornsby *et al*., 2015). A detailed description of library-F design is included in Section 5.1.2. The commercial availability of high-quality synthetic genes and oligonucleotides has played a crucial role in the construction of synthetic antibody libraries. In particular, the development of trinucleotide phosphoramidite cassette technology and has enabled the synthesis of mutagenic oligonucleotides that can encode desired proportions of tailored amino acid diversities at each random position (Miersch and Sidhu, 2012; Shim 2016).

Phage-displayed single-framework synthetic antibody libraries are constructed in three major steps. First, the chosen antibody framework is assembled from synthetic gene fragments and sub-cloned into a phagemid vector. Second, CDR regions within the template phagemid are diversified using mutagenic oligonucleotides. The incorporation of mutagenic oligonucleotides

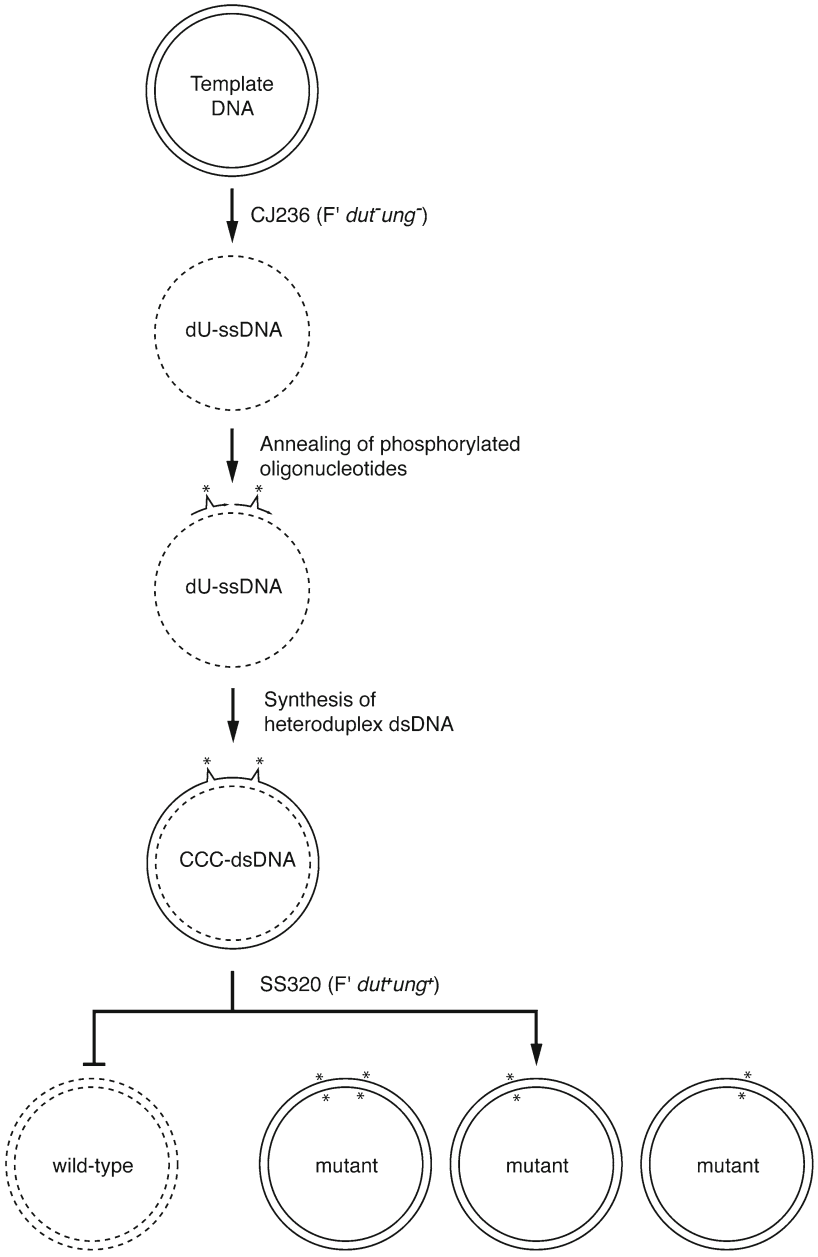


Figure 2.6: Library construction by Kunkel mutagenesis. Uracil-containing single-stranded DNA form (dU-ssDNA) of the template phagemid is prepared using a *dut*$^{-}$/*ung*$^{-}$ *E. coli* host (CJ236). Phosphorylated mutagenic oligonucleotides (arrows) designed to introduce mutations (asterisk) in the CDRs of interest are annealed to the dU-ssDNA template strand. A covalently closed, circular, double-stranded DNA (CCC-dsDNA) heteroduplex is synthesized *in vitro* by T7 DNA polymerase and T4 DNA ligase in the presence of dNTPs and ATP. The resulting CCC-dsDNA is then electroporated into a *dut*$^{+}$/*ung*$^{+}$ *E. coli* host (SS320) that inactivates the uracil-containing wild-type strand (dashed circle) leading to the preferential amplification of mutated DNA. Figure 2.6 reprinted with permission from Springer (Chen and Sidhu, 2014).

into the phagemid vector is typically accomplished using either cassette-based methods with restriction enzyme cloning or site-specific annealing to ssDNA followed by whole phagemid synthesis. Third, the phagemid library is introduced into *E. coli* cells by electroporation followed by helper-phage infection and library phage production (Miersch and Sidhu, 2012; Fellouse and Pal, 2015). In this work, we started with a pHP153 phagemid that contains the Hu4D5-8 Fab framework (Persson *et al*., 2013). To incorporate mutagenic oligonucleotides to the CDR regions of the phagemid, we chose to use the classical oligonucleotide-directed mutagenesis method of Kunkel *et al*. (1987). First, uracil-containing template phagemid ssDNA is prepared using a *dut*$^{-}$/*ung*$^{-}$ *E. coli* host. Second, mutagenic oligonucleotides are annealed to the template ssDNA to prime the polymerase-dependent *in vitro* synthesis of complementary DNA strand. The whole phagemid vector thus synthesized is ligated to form a covalently closed, circular, double-stranded DNA (CCC-dsDNA) heteroduplex. Third, CCC-dsDNA is electroporated into a *dut*$^{+}$/*ung*$^{+}$ *E. coli* host that inactivates the uracil-containing template DNA leading to the preferential amplification of mutated DNA (Fellouse and Sidhu, 2006; Adams *et al*., 2014). Kunkel mutagenesis does not require prior assembly of overlapping PCR fragments or restriction enzyme sites, enables simultaneous diversification of multiple CDRs in a single mutagenesis reaction, and allows using pools of mutagenic oligonucleotides of varying lengths thus facilitating the incorporation of length diversity into CDRs (Miersch and Sidhu, 2012). Upon electroporation of CCC-dsDNA into *E. coli*, the phagemid replicates as a dsDNA. Helper phage infection initiates ssDNA replication and provides the necessary proteins and viral machinery for the production of mature phage particles displaying Fabs encoded in phagemid vectors (Fellouse and Sidhu, 2006; Adams *et al*., 2014).

Phage display panning is the process of screening phage-displayed libraries for isolating binders against a given target. Typically, selections are conducted against purified target proteins immobilized onto a solid support. The selection process aims to sequentially enrich the clones of the phage-displayed antibody library, which bind to the target of interest as the library undergoes successive rounds of selection. At each selection round, negative selections are conducted to remove phage clones that bind to undesired regions of the target or any other control proteins. Phage pools are then exposed to the immobilized target, non-binding phages are washed away and phages retained by the target are eluted and amplified for the next round of selection. After each selection round, the enrichment in the selection process is measured by comparing the

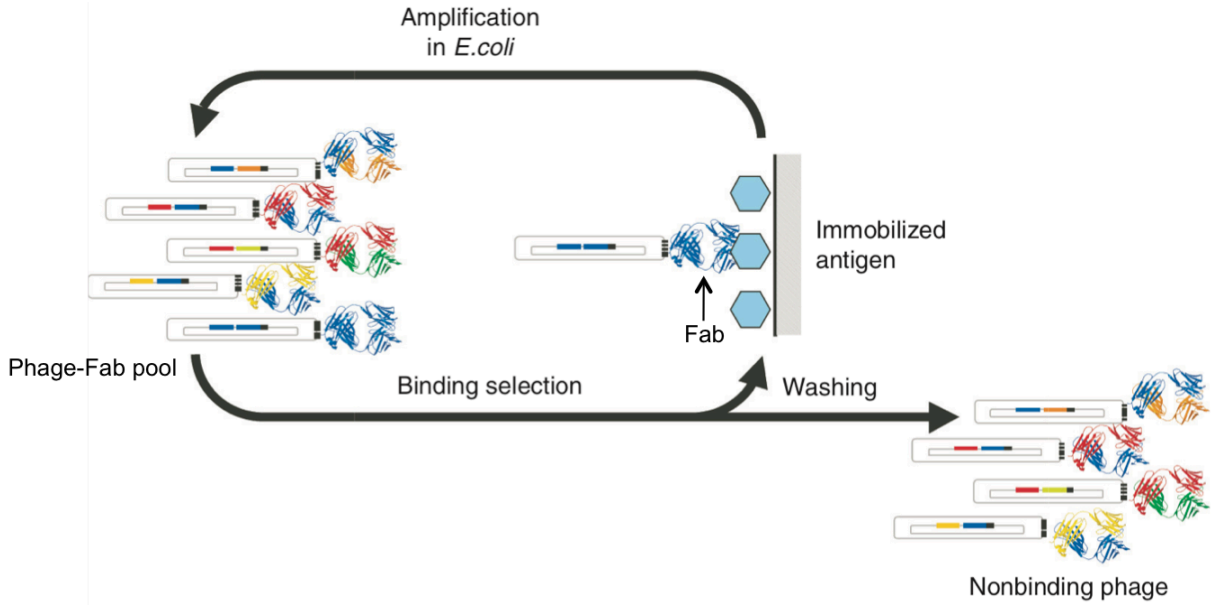


Figure 2.7: Phage display selections of Fabs. Fab libraries are displayed on the surfaces of M13 phage particles as fusions to the PIII phage coat protein. Each phage particle displays a unique Fab protein and also encapsulates a phagemid that contains the Fab-encoding DNA. The phage-displayed Fab library is incubated with the immobilized antigen, and non-binding phages are washed away. Antigen-binding phages retained by the immobilized antigen are amplified in an *E. coli* host. The amplified pool is used for additional rounds of selection to eventually obtain a phage population that is dominated by antigen-binding clones. At this point, Fab sequences are decoded by subjecting the phage pools and/or individual phage clones to DNA sequencing. Figure 2.7 reprinted with permission from Nature Publishing Group (Sidhu and Fellouse, 2006).

phage titers eluted from the target protein and control proteins. After a few rounds of binding selection and amplification, target-binding antibody fragments are isolated from the phage pool and their identities are decoded using Sanger sequencing (Reviewed by Hoogenboom, 2005; Sidhu and Fellouse, 2006; Bradbury *et al*., 2011; Chan *et al*., 2014).

2.3 Novel Strategies for Discovering Antibodies

2.3.1 Phage Display Challenges and Selection Strategies

The success of a selection experiment depends on the quality of the phage-displayed antibody library, the immobilized target and the phagemid/phage/host system. During library design and construction, quality of the antibody library may be affected due to various factors such as less-stable antibody frameworks, undesirable framework changes, CDRs designs that affect folding and stability of Fabs, inefficient library mutagenesis reactions, and sub-optimal phagemid systems that affect phage assembly and Fab display. As a result, the library may not

contain sufficient functional diversity to generate antibodies against a broad range of antigens (Ponsel *et al*., 2011; Zhai *et al*., 2011; Shim 2015). A properly folded and immobilized antigen is also critical in determining the success of a selection experiment. Immobilization of proteins on plates or tubes may lead to some denaturation thus reducing the proportion of correctly folded antigen, result in non-uniform presentation of proteins, disrupt the functional or naïve conformation and block access to relevant epitopes on the target (Koide *et al*., 2009; Miersch and Sidhu, 2012; Jara-Acevedo *et al*., 2016). Finally, antibody selection by phage display is not only driven by binding events, but also the amplification phase of panning has a significant effect on the enrichment of antibody fragments during successive rounds of selection. The amplification of individual clones in the library is not uniform because some phage-displayed antibody fragments can influence the infection, propagation and production of M13 and *E. coli*. While phagemids, antibody libraries and panning protocols have been optimized to minimize the amplification bias, an entirely affinity-driven selection is hard to achieve due to the intrinsic limitations of phage display that relies on M13 and *E. coli* biology (Derda *et al*., 2011; Saggy *et al*., 2012; Naso *et al*., 2014; Shim, 2015).

Phage display screening of antibody libraries against immobilized targets does not give rise to antibody fragments in all selection experiments. Further, antibody fragments obtained from a successful selection may not possess favorable biophysical, biochemical, cell biological, pharmacological and clinical properties. Therefore, novel strategies are required for increasing the success rate of phage display selections and for obtaining useful antibodies with desirable properties. We have classified these strategies into six groups: (1) novel library designs; (2) target modification; (3) manipulating the selection environment; (4) isolation of rare clones from selection outputs; (5) post-selection strategies; and (6) post-clone rescue strategies. Novel library designs include either next-generation synthetic antibody libraries with improved biophysical properties, favorable production characteristics and reduced immunogenicity (Ponsel *et al*., 2011; Tiller *et al*., 2013) or motif-grafted synthetic antibody libraries that are designed to direct the antibody library towards a specific region on a protein (Tiller and Tessier, 2015; Miersch *et al*., 2017). If target immobilization by passive adsorption is unsuccessful, affinity tags can be engineered into the target protein for immobilization into magnetic beads. Also, selections can be conducted against complex surfaces such as normal and tumorigenic cells, engineered bacterial, yeast and mammalian cells, tumor histological samples and proteoliposomes (Miersch and

Sidhu, 2012; Jara-Acevedo *et al*., 2016). In phage display, the selection conditions can be tailored or controlled by addition of ligands or competitors, decreasing target concentrations, extending washing steps, increasing temperature, including negative selections and competitive elution techniques. These strategies increase the chances of obtaining clones with desired binding properties including high/low affinity, high/low specificity, thermo-stability, regional or conformational specificity and species cross-reactivity (Paduch *et al*., 2013; Dennis, 2015). The use of high-throughput approaches and robotic liquid-handling devices for phage display screening can increase the number of unique antibody clones isolated from selection outputs (Hornsby *et al*., 2015). Also, phage pools can be subjected to next-generation sequencing (NGS) analysis for identifying and reconstructing low-frequency rare clones from selection outputs (D'Angelo *et al*., 2014). Following one or two rounds of selection, the CDR regions in phage pools can be subjected to additional rounds of mutagenesis or sequence diversity can be sub-cloned into alternate display platforms before resuming further rounds of selection (Ferrara *et al*., 2012). Also, NGS analysis of selection pools can provide insight for constructing target-specific focused libraries, which can be screened for isolating binders with improved properties (Ravn *et al*., 2010; Mathonet and Ullman, 2013; Koenig *et al*., 2015). Once an antibody fragment has been identified and its sequence decoded, second-generation libraries can be constructed and screened for modifying its properties. These libraries, commonly referred to as affinity maturation libraries, are constructed by soft randomization of CDRs or by incorporating restricted or tailored diversity to CDRs. Knowledge obtained from CDR homology scanning, CDR alanine scanning or antibody-antigen structures can be used for designing randomization schemes in secondary libraries or for engineering rational mutations into CDRs (Miersch and Sidhu, 2012; Marvin and Lowman, 2015). Site-directed mutations can be engineered to Fab or Fc framework regions based on previous observations or crystal structures to increase the stability or alter the effector function of antibodies (Beck *et al*., 2010).

2.3.2 Next-Generation Sequencing in Antibody Discovery

In a typical antibody selection experiment, once the library is panned against the target, antigen-specific antibody fragments must be recovered from phage selection pools. To accomplish this, hundreds to thousands of colonies from phage selection outputs are first interrogated by binding assays (phage/Fab ELISA) in 96/384 well plates, and then identities of ELISA-positive clones are determined by Sanger sequencing (Ravn *et al*., 2010). In phage

display, many binders present in earlier selection rounds do not get enriched and remain at very low frequencies in later selection rounds. The iterative nature of panning experiments (rounds of binding selection and amplification) often enrich for binders that exhibit certain growth advantages in *E. coli* (Derda *et al*., 2011; Saggy *et al*., 2012). The conventional clone recovery method (screening followed by sequencing) leads to the repeated identification of the same enriched clones that possessed growth-advantage characteristics in *E. coli*. With this method, it is not possible to recover Fab clones that are present at low frequencies in phage pools (termed as rare clones). If we were to switch the order of this clone recovery process (sequencing followed by screening), we could avoid the repeated identification of growth-advantaged, high frequency binders. In addition, this new approach can significantly reduce the amount of work required during the upfront screening process, and can help to identify and recover many unique clones that are present from high to very-low frequencies in phage pools (Naso *et al*., 2014; Sasso *et al*., 2015; Lopez *et al*., 2017).

Sanger sequencing is used to decode the sequences of Fab-encoding phagemids recovered from phage pools. It requires the isolation of individual phagemids from phage pools, and can only sequence a few hundred clones routinely. Since each selection round in phage display gives an output of up to ~10^6 sequences, only a small fraction of the sequence diversity can be sampled by Sanger sequencing. Though Sanger sequencing is sufficient for the typical clone recovery process (screening followed by sequencing), the use of next-generation sequencing (NGS) becomes necessary if one has to capture and interrogate the entire sequence diversity present in phage display selection outcomes. With NGS, entire phage pools can be subjected to sequencing (isolation of individual phagemids is not required), and up to ~10^6 – 10^9 individual sequences can be obtained routinely (Fischer, 2011; Glanville *et al*., 2015). Up to ten NGS platforms exist each with its own advantages and preferred applications, with primary variables being read length, data quality and quantity, time and cost (Metzger, 2010; Loman *et al*., 2012). Three commonly used NGS platforms in the antibody field include 454 (Roche), MiSeq (Illumina) and Ion Torrent (Life Technologies).

The Fischer lab at Novimmune was the first to use NGS in antibody phage display. The Illumina platform was used to sequence the CDRH3 region of a multiple-framework synthetic scFv library. NGS analysis was used to assess the quality of the library and to follow changes in heavy chain germline usage and CDRH3 length distribution over three rounds of selection. A

Table 2.3: Comparison of commonly used next-generation sequencing platforms

Platform	Type of Seq.	Max read length (bp)	Throughput	Cost (lowest)	Accuracy	Time	Type of error
MiSeq (Illumina) V2/V3	2×300	600	25×10^6/lane	$1750/lane	>70% reads at 99.9%	55h	Substitution
	2×150	300	16×10^6/lane	$1100/lane	>80% reads at 99.9%	24h	
IonTorrent (LifeTech)-318	1×400	400	5.5×10^6/chip	$749/chip	>99%	5h	Insertion Deletion
	1×200	200				3h	
454 (Roche)-GS-FLWX+	1×700	700	50,000 in 1/8 plate	$2400/ 1/8plate	99.99%	23h	Insertion Deletion
	1×450	450		$1900/ 1/8plate		10h	

Type of sequencing, maximum read length, number of reads, reagent costs, time taken, error rate and types of error for three commonly used NGS platforms. Table modified from Glanville *et al*., 2015.

strategy was also developed to rescue rare scFv clones from phage pools using fragment assembly (Ravn *et al*., 2010). The Fischer lab used the same approach to sequence the CDRH3 region of semi-synthetic scFv libraries, and to rescue rare scFv clones from phage selection outputs (Venet *et al*., 2012; Ravn *et al*., 2013). The Lerner lab used Roche's 454 to sequence the V_H region of a phage selection output originated from a natural scFv library. Three strategies were tested for recovering scFv clones based on CDRH3 sequences: fragment assembly, rolling circle amplification, and hybridization using biotin probes (Zhang *et al*., 2011). A few labs have used Roche's 454 to sequence the V_H and V_L regions of synthetic Fab/scFv libraries, and NGS information was used to assess the quality of libraries (Zhai *et al*., 2011; Tiller *et al*., 2013; Mahon *et al*., 2013). The Bradbury lab used the Ion Torrent platform to sequence the CDRH3 region of two different selection outputs originated from a natural scFv library. A rescue strategy was also developed to isolate scFv clones from selection outputs using inverse PCR and ligation

(D'Angelo *et al*., 2014; Spiliotopoulos *et al*., 2015). Recently, MiSeq (Illumina) was used to sequence the CDRH2-CDRH3 region of a selection output originated from a single-framework synthetic Fab library (Lopez *et al*., 2017), and the V_H region of selection outputs originated from immune scFv libraries (Yang *et al*., 2017). Both groups used CDRH3 information for analyzing the amino acid composition, for monitoring the enrichment of clones during the selection process, and for retrieving rare Fab/scFv clones by fragment assembly (Lopez *et al*., 2017; Yang *et al*., 2017).

For antibody engineering, a key aspect is to obtain the entire sequence of highly diverse V_H and V_L chains. At present, none of the NGS platforms can provide sufficient read lengths to sequence the full-length Fab genes. Roche's 454 offers the longest read length (700 bp), which is only sufficient to cover one of the variable chains (DeKosky *et al*., 2014). Due to this read-length limitation in short-read DNA sequencing platforms, NGS analysis of phage-displayed Fab/scFv libraries or selection outputs is usually restricted to CDRH3. It is worth mentioning here that CDRH3 is typically the most heavily diversified CDR in Fab/scFv libraries due to its dominant role in antigen recognition. CDRH3 sequencing has been used to assess the quality of libraries, to monitor the evolution of Fabs during selections and to characterize the changes in CDR length or amino acid distribution during selections (Ravn *et al*., 2010; D'Angelo *et al*., 2014). Reconstruction of rare clones, however, is not the same as sequence identification. To reconstruct full length Fab or scFv clones from NGS information, CDRH3 information alone is not sufficient, as antibody libraries typically contain two, four or all six diversified CDRs. To circumvent this, a few strategies have been tested to rescue Fab/scFv clones from selection outputs based only on CDRH3 information. These strategies use both hybridization- and PCR-based cloning techniques to rescue the entire sequences of Fab/scFv clones (Ravn *et al*., 2010; Zhang *et al*., 2011; D'Angelo *et al*., 2014; Lopez *et al*., 2017; Yang *et al*., 2017). Gene synthesis has also been used to reconstruct scFv clones from NGS information (Lövgren *et al*., 2016). Rescued Fab/scFv clones had been shown to interact with their targets by ELISA or flow cytometry (Ravn *et al*., 2010; D'Angelo *et al*., 2014; Lopez *et al*., 2017; Yang *et al*., 2017). In this work, we describe a rapid and simple method for reconstructing rare Fab clones from NGS information, and also compare the binding properties of reconstructed rare clones with high-frequency top clones.

2.3.3 Motif-Grafting in Antibody Discovery

Protein-protein interactions form the molecular basis of most physiological and pathological processes. From a structural standpoint, protein-protein interactions can be considered as interactions between two epitopes. Epitopes are divided into two categories: linear or continuous epitopes and discontinuous epitopes (Van-Regenmortel, 2009). Linear or continuous epitopes are formed by a single contiguous stretch of amino acids; however, it must be noted that not necessarily each residue in this peptide segment interacts with the partner. The peptide segment can exist as a single secondary structure element (β-strands or turns) or with no clear secondary structures (patches). Discontinuous epitopes are formed by multiple secondary structure elements (α-helixes, β-strands or turns) and complex patches originating from different regions of a binding partner (Jubb *et al*., 2012; Nero *et al*., 2014; Scott *et al*., 2016). An antigen-antibody interaction is essentially a protein-protein interaction, which involves interaction between the epitope on the antigen and the antigen-binding region on the antibody. The antigen-binding region on the antibody is referred to as the "paratope", which is formed by one to six CDR loops and selective framework residues (Sundberg, 2009).

Motif grafting is one of the key approaches used for rationally designing antibodies based solely on the sequence and/or structure of the antigen. Motif grafting relies on engineering protein interaction epitopes into CDRs for designing antibodies with desired target specificity by mimicking natural protein interactions (Tiller and Tessier, 2015). Many linear peptides or discrete motifs have been successfully grafted into CDRs for generating antibodies with desirable binding and functional properties (McLane *et al*., 1995; Moroncini *et al*., 2004; Simon *et al*., 2005; Kogelberg *et al*., 2008; Perchiacca *et al*., 2012; Ladiwala *et al*., 2012). Recent studies have shown that small protein domains can also be grafted into CDRs of certain antibody frameworks to obtain fully functional antibodies (Zhang *et al*., 2013; Zhang *et al*., 2014; Zhang *et al*., 2015; Liu *et al*., 2015A; Liu *et al*., 2016). A few studies have combined motif grafting with combinatorial approaches to improve the binding properties of grafted antibodies. Amino acids flanking the graft (Barbas *et al*., 1993; Frederickson *et al*., 2006; Peng *et al*., 2015; Lee *et al*., 2015) or amino acids within ungrafted CDRs (van den Beucken *et al*., 2001; Koerber *et al*., 2013) have been optimized using directed evolution to obtain grafted antibody variants with high affinity and specificity. These studies highlight the use of motif grafting approaches in rational and semi-rational design of synthetic antibodies.

The use of antibodies as support scaffolds to mimic protein-protein interactions provides many advantages: (1) the antibody framework provides a good platform with respect to stability, rigidity, mobility, flexibility, cooperativity and interactivity; (2) CDRs of antibodies show a high level of plasticity by exhibiting amino acid, length and conformational diversity, and are able to tolerate grafts of any reasonable size or character from a variety of sources; (3) antibodies are compatible with combinatorial diversification and selection methods for further optimization of binding properties; (4) reliable information on molecular, structural, functional, pharmacological and clinical properties of full-length antibodies and antibody fragments is available for design and development of epitope mimics; and (5) antibodies can be designed with desirable biochemical, biophysical and pharmacokinetic properties and favorable production characteristics.

The antibody paratope is discontinuous; it is composed of multiple CDR loops that are spatially brought together by the antibody framework. The interaction of multiple secondary structures with the antigen enhances the affinity and specificity of binding. The presence of multiple CDRs on a single molecular scaffold enables antibodies to be used as discontinuous epitope mimics. In discontinuous epitopes, the shape, conformation and positioning of different peptide segments with respect to each other are crucial for mediating an interaction. The exact mimicry of these features on antibody CDRs can be extremely difficult (Werkhoven and Liskamp, 2013). For this reason, a strategy is proposed in this work for partially resembling discontinuous epitopes using antibody CDRs. First, a secondary structure element central for the interaction is identified and grafted into CDRH3 of a Fab framework. Second, to compensate for other interactions in the protein-protein interface, amino acids within ungrafted CDRs are diversified and selected for optimal binding by combinatorial phage display. This approach could be used for semi-rationally designing antibody libraries that are biased towards binding to a specific region on the protein target for disrupting a discontinuous protein-protein interaction.

2.4 Background Information about Protein Targets used in this PhD Project

2.4.1 Notch and Jagged Receptors

Notch signaling is a highly-conserved pathway that influences multiple cell fate decisions including, proliferation, apoptosis, differentiation, migration and angiogenesis in developing and adult metazoan organisms. In mammals, Notch signaling is initiated by four Notch receptors

(Notch1-4) and five Delta/Serrate/LAG-2 (DSL) ligands (Jagged-1/2 and DLL-1/3/4), all of which are modular, type-I single-pass transmembrane proteins. The extracellular region of Notch contains a series of epidermal growth factor (EGF)-like repeats that are required for ligand binding followed by a negative regulatory region and a hetero-dimerization domain. The extracellular region of DSL ligands contains a MNNL domain (module at the N-terminus of Notch ligands) followed by a DSL domain and a series of EGF-like repeats like Notch receptors. The intracellular region of Notch contains numerous well-defined domains whereas the intracellular region is not conserved in DSL ligands (reviewed by Gordon *et al*., 2008 and Espinoza *et al*., 2013).

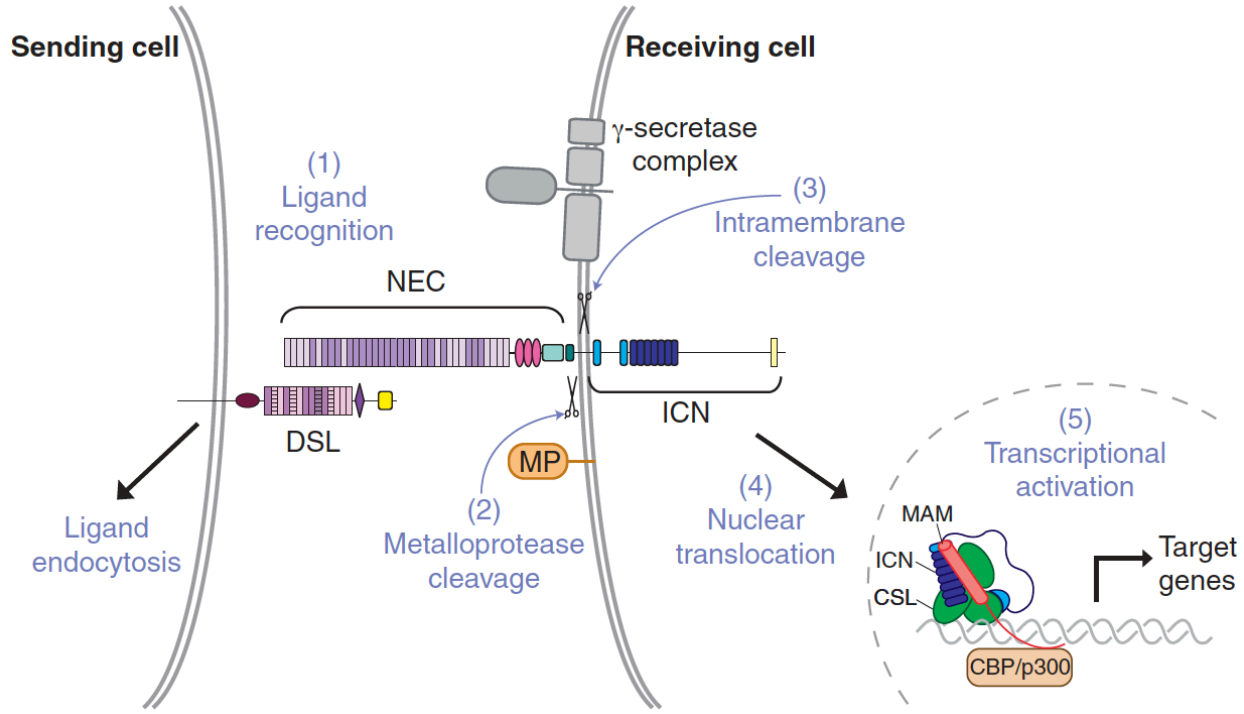


Figure 2.8: Key steps in the Notch signaling pathway: (1) Interaction between a DSL ligand on the signal-sending cell and a Notch receptor on the signal-receiving cell; **(2)** Cleavage of Notch at site S2 within the juxtamembrane domain by a metalloprotease of the ADAM family; **(3)** Cleavage of Notch at site S3 within the transmembrane domain by the γ-secretase protein complex; **(4)** Release of intracellular Notch from the membrane and its translocation to the nucleus; and **(5)** Formation of a transcriptional activation complex with CSL and MAM, thus inducing the transcription of target genes. (NEC: Notch extracellular; DSL: Delta/Serrate/LAG-2; ICN: Intracellular Notch; MAM: Mastermind; CSL: C-promoter-binding factor; CBP/p300: CREB-binding protein/ E1A binding protein p300). Figure 2.8 reprinted with permission from Company of Biologists (Gordon *et al*., 2008).

The canonical short-range Notch signaling happens between two adjacent cells, with the signal-sending cell expressing the DSL ligand and signal-receiving cell expressing the Notch receptor. Bi-molecular interaction between the Notch receptor and the DSL ligand at the cell surfaces initiates a process called regulated intramembrane proteolysis, in which the Notch receptor is cleaved within the extracellular hetero-dimerization domain by a metalloprotease of the ADAM family. This DSL-dependent metalloprotease processing renders Notch sensitive to the γ-secretase protein complex. γ-Secretase cleaves the Notch receptor within the transmembrane region and releases the Notch intracellular domain. The intracellular domain translocates into the nucleus where it forms a transcriptional activation complex and regulates Notch-responsive genes (reviewed by Gordon *et al*., 2008 and Wang, 2011).

In addition to its physiological roles, deregulation of Notch signaling is associated with developmental disorders, neurological diseases, solid cancers and hematologic malignancies. These pathological conditions result from overexpression or gain/loss-of function mutations within Notch and Jagged receptors (Ranganathan *et al*., 2011). Two kinds of inhibitors have been developed to inhibit Notch signaling: small molecules targeting the γ-secretase complex and monoclonal antibodies targeting Notch or Jagged ectodomains. Since γ-secretase inhibitors inhibit the proteolysis of multiple transmembrane proteins including all four Notch receptors, they caused severe toxic side effects in clinical trials. Similar side effects were also observed with pan-Notch specific monoclonal antibodies (Takebe *et al*., 2014). Synthetic antibodies targeting the negative-regulatory region of Notch-1 and receptor-binding region of Jagged-1 bound specifically to Notch and Jagged receptors, respectively. They also exhibited potent and selective inhibition of Notch-1 or Jagged-1 signaling in tumor models and showed promising results in pre-clinical studies (Wu *et al*., 2010 and Lafkas et al., 2015). Thus, there is a great interest in developing paralogue-specific synthetic antibodies against Notch receptors and DSL ligands.

2.4.2 Epidermal Growth Factor Receptor

The epidermal growth factor (EGF) family of receptor tyrosine kinases (also known as ErbB or HER) plays a well-established role in the growth, differentiation, migration and survival of normal adult tissues (Yarden and Sliwkowski, 2001). This family is comprised of four closely related receptors including EGFR (ErbB1/HER1), HER2 (EGFR2/ErbB2), HER3 (ErbB3) and HER4 (ErbB4), all of which share several structural and functional features including a 4-domain

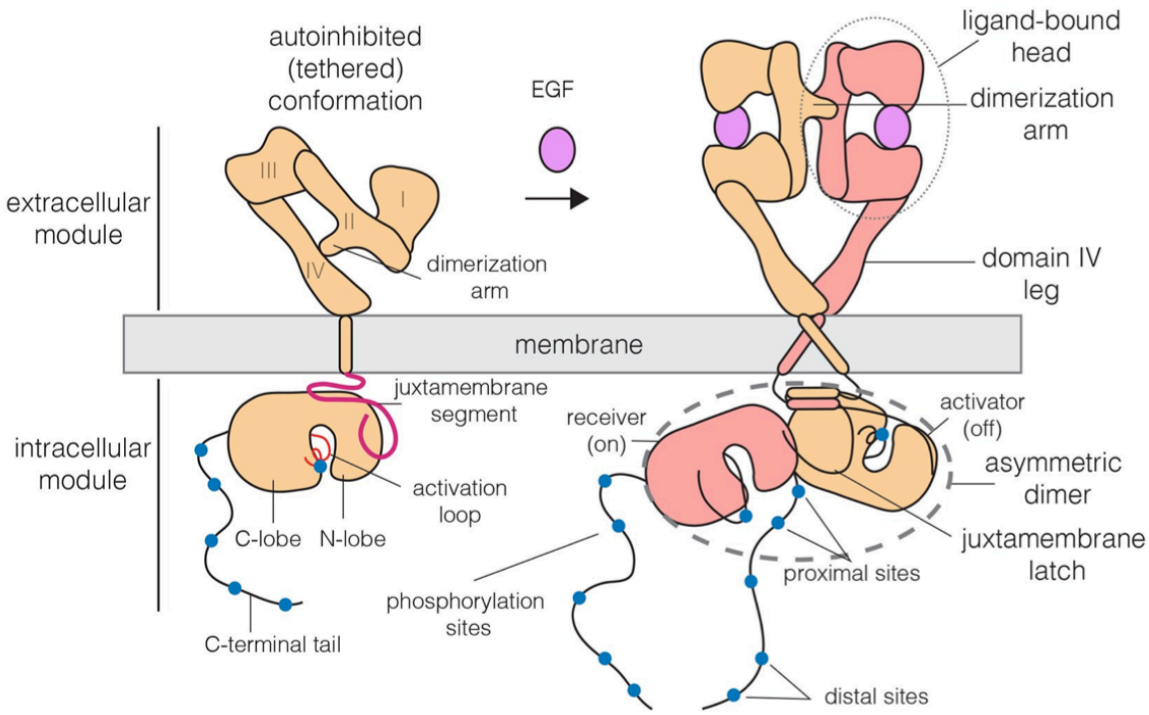


Figure 2.9: Mechanism of EGF-induced EGFR activation. The extracellular module of EGFR contains four domains (I to IV, distal to membrane proximal). Domains I and III form the ligand-binding site on the receptor. Domain II contains the 'dimerization arm'. The intracellular module of EGFR contains a tyrosine kinase domain and a C-terminal tail. In the absence of EGF, the EGFR monomer predominantly exists in an auto-inhibited conformation, where domains II and IV form an intramolecular interaction. In the presence of EGF, EGFR adopts an extended conformation, where domain II is exposed for promoting receptor dimerization. EGFR dimerization activates the intracellular kinase domain of EGFR, resulting in autophosphorylation of the C-terminal tail and initiation of downstream signaling pathways. Figure 2.9 reprinted under the terms of the Creative Commons license (Huang *et al*., 2016).

(I to IV, distal to membrane proximal) modular extracellular region, a single transmembrane domain, and intracellular tyrosine kinase and regulatory domains (Ferguson, 2008; Lemmon *et al*., 2014). In the absence of a ligand, EGFR predominantly exists in a tethered conformation in which domains II and IV interact and impede ligand binding. Upon ligand binding to domains I and III, EGFR adopts an extended conformation in which domain II is exposed for promoting receptor dimerization (Ferguson, 2008; Bessman *et al*., 2014). The domain II dimer interface is composed of multiple discontinuous contacts, and a β-hairpin loop in domain II, referred to as the dimerization arm has been shown to be crucial for dimerization (Ogiso *et al*., 2002; Garrett *et al*., 2002; Burgess *et al*., 2003). Receptor dimerization stabilizes the extended conformation of EGFR and triggers phosphorylation events in the cytoplasmic tail of the receptors. These events

enable docking of intracellular proteins bearing SH2 domains to phospho-residues, resulting in the transmission of signals via the MAPK, PI3K and STAT pathways (Ferguson, 2008; Bessman *et al*., 2014; Huang *et al*., 2016).

In addition to their physiological roles, ErbB receptor deregulation and aberrant expression are commonly observed features of cancers forming solid tumors (Wang *et al*., 2011; Huang *et al*., 2012). EGFR overexpression in particular was confirmed to transform normal fibroblast cells into tumorigenic cells long ago, and the frequent overexpression of EGFR in many cancers and the ensuing correlation with prognosis indicates that EGFR plays a central role in oncogenesis, progression and metastasis (Hudziak *et al*., 1987; Velu *et al*., 1987). Increased EGFR expression in the absence of activating mutations has been confirmed in a wide variety of cancers including head and neck, esophageal, lung, colorectal, ovarian, prostate, and breast. EGFR overexpression is now known to be a primary driver not only for tumor initiation, but also metastasis and invasion, promoting the formation of secondary tumors. Based on the correlation between high levels of EGFR expression in solid tumors and disease progression and prognosis, EGFR has received substantial attention as an anti-cancer target, which has resulted in the development of a host of drugs aimed at inhibiting aberrant EGFR-dependent signals that drive tumor development (reviewed by Yewale *et al*., 2013; Tebbutt *et al*., 2013; Arteaga and Engelman, 2014).

Currently two classes of anti-EGFR drugs exist and are approved for the treatment of cancers: Small molecule tyrosine kinase inhibitors (gefitinib and erlotinib) and monoclonal antibodies (cetuximab, panitumumab and nimotuzumab). These drugs have shown promise as mono-, combination-, and follow-up therapies in the treatment of colorectal, non-small cell lung, pancreatic, and breast cancer as well as squamous cell carcinoma of the head and neck. Although patients have derived clinical benefit from small molecule tyrosine kinase inhibitors, monoclonal antibody therapies are attractive as antibodies can mediate pleiotropic activities that inhibit receptor kinase complex activity and reduce tumor burden by (1) blocking ligand binding and the active receptor conformation, (2) inhibition of receptor dimerization, (3) recruitment of immune effectors that mediate cell killing, (4) receptor internalization and down-regulation, and (5) inhibition of ectodomain proteolysis (reviewed by Peipp *et al*., 2008; Fauvel and Yasri, 2014; Arteaga and Engelman, 2014).

One of the main disadvantages of current EGFR-directed therapies is the potential for primary resistance (in which individuals initially do not respond to drug) or secondary resistance (in which resistance to drug is acquired upon extended exposure to drug), both of which have limited the efficacy and utility of anti-EGFR drugs and have resulted in reports of poor correlation between EGFR expression and drug responses. Attempts to characterize resistance mechanisms have found several likely causes including receptor overexpression and trans-activation of alternate ErbB receptor family members, as well as mutations in downstream effectors of the EGFR pathway including MAPK (BRAF) and PI3K (KRAS, PTEN) pathways, which may partially decouple signaling from the receptor (reviewed by Kruser and Wheeler, 2010; Wheeler *et al*., 2010; Camidge *et al*., 2014). Clearly, there is still substantial need for novel therapeutics that targets this receptor with unique mechanisms that can help to overcome resistance. Synthetic antibodies are ideally suited to this task as they can be designed to target unique epitopes and are a well-tolerated, routinely engineered class of biologics.

3. OBJECTIVE, HYPOTHESES AND SPECIFIC AIMS

The **objective** of this PhD project is to develop and validate methods for advancing the applications of two techniques, NGS and motif grafting, in synthetic antibody discovery.

A few strategies have been described to rescue low-frequency Fab/scFv clones from phage selection outputs using NGS information. These strategies involve gene synthesis, fragment assembly, multi-step PCR reactions and ligation procedures, which are time consuming and expensive. These methods are relatively low-throughput compared with the high-throughput of NGS and phage display selections. Further, none of these strategies eliminated the use of Sanger sequencing during clone rescue or reconstruction. The **objective** of this NGS project is to develop a method for reconstructing rare Fab clones from NGS information, which overcomes all of afore-mentioned limitations. We **hypothesize** that rare Fab clones possessing favorable binding properties can be reconstructed from NGS information in a straightforward manner. **Specific aims** for this project include: **(1)** develop methods for sequencing phage-Fab selection outputs and to reconstruct Fabs from NGS information; **(2)** design, construct and validate a new single-framework synthetic Fab library for NGS-assisted selections; and **(3)** compare binding properties of Fabs isolated using conventional and NGS approaches.

Previous studies have shown that linear peptides, discrete motifs or small domains can be grafted into CDRs to design antibodies with desired binding properties. A few studies have combined motif grafting with combinatorial approaches to improve the binding properties of grafted antibodies. The **objective** of this motif-grafting project is to test whether structural information on discontinuous protein interfaces could be used for semi-rationally designing synthetic Fab libraries that are biased towards interacting with a desired site on the target receptor. We designed and constructed a structure-guided Fab library by grafting a well-defined EGFR interaction motif into CDRH3 and by diversifying few other CDRs of a Fab framework. We **hypothesize** that the structure-guided Fab library can be used for isolating Fabs that recognize the native interaction domain in the target receptor. **Specific aims** for this project include: **(1)** design, construct and screen an EGFR structure-guided synthetic Fab library; **(2)** test the affinity, specificity and activity of anti-EGFR Fabs isolated from the structure-guided and naïve synthetic Fab libraries; and **(3)** characterize the epitopes of anti-EGFR Fabs.

4. MATERIALS AND METHODS

4.1 General Molecular Biology and Microbiology Protocols

Plasmid DNA was extracted and purified from *E. coli* using plasmid mini-preparation and maxi-preparation kits (Qiagen/Bio-Basic). PCR products and restriction enzyme digested plasmids were purified using PCR cleanup kits (Qiagen/Bio-Basic) and gel purification kits (Qiagen/Bio-Basic), respectively. Single-stranded DNA was extracted and purified from M13 phage solution using the Spin M13 kit (Qiagen). Agarose gel electrophoresis and sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) were used to separate DNA and proteins based on their molecular weight, respectively. Plasmids were sequenced on a 3500 Series Genetic Analyzer (Applied Biosystems) using the BigDye Terminator 3.1 Cycle Sequencing Kit (Applied Biosystems).

Bacterial strains were cultured and propagated using standard microbiology procedures. 2YT broth and 2YT agar were used to cultivate *E. coli*. The following *E. coli* strains were used in this work: DH10B and MC1061 for general cloning purposes and plasmid amplification, CJ236 for dU-ssDNA production, SS320 and SR320 for library generation and phage production, XL1 Blue and 5-α $F'I^q$ for bacteriophage amplification, and BL21 for Fab expression and purification.

Electrocompetent DH10B and MC1061 cells were prepared according to standard protocols (Ausubel *et al*., 1988). To transform DH10B and MC1061 cells, 1 μL of plasmid (50-150 ng) was mixed with 50 μL of electrocompetent cells, transferred to an electroporation cuvette, and the cuvette was subjected to a short electric pulse using a gene pulser (Bio-Rad). The cells were recovered immediately, transferred into 1 mL of pre-warmed SOC media and incubated for 30 min at 37 °C. The cells were harvested and plated onto 2YT agar plates containing the appropriate antibiotic for plasmid selection.

4.2 Kunkel Mutagenesis

Kunkel mutagenesis was used to: (1) introduce site-directed mutations into the antibody framework or CDRs in the pHP153 phagemid; (2) diversify one or multiple CDRs in the pHP153 phagemid for generating Fab libraries; (3) create EGFR mutants and domain truncations in the EGFR expression plasmid for epitope mapping studies; and (4) delete the framework regions between diversified CDRs in phage pools for reducing the NGS amplicon length. For the first

three purposes, we used the conventional mutagenesis approach in which dU-ssDNA is prepared in *dut*$^{-}$/*ung*$^{-}$ *E. coli*, dU-ssDNA is converted to CCC-DNA *in vitro*, and CCC-dsDNA is electroporated into a *dut*$^{+}$/*ung*$^{+}$ *E. coli* host. Kunkel mutagenesis (Kunkel *et al*., 1987) was carried out using protocols established by the Sidhu lab (Fellouse and Sidhu, 2006; Tonikian *et al*., 2007; Nelson and Sidhu, 2012). The method used for NGS amplicon preparation is explained in Section 4.8.

To prepare dU-ssDNA, *dut*$^{-}$/*ung*$^{-}$ CJ236 *E. coli* cells harboring the appropriate phagemid were grown, infected with M13KO7 helper phage, and incubated in 2YT media supplemented with 100 μg/mL carbenicillin, 25 μg/mL kanamycin, and 0.25 μg/mL uridine for 20 hrs at 37 °C with shaking at 200 rpm. Phage particles were precipitated from the culture supernatant, and dU-ssDNA was extracted and purified from phage particles using the Spin M13 kit (Qiagen). The dU-ssDNA template was converted to CCC-dsDNA in three steps: phosphorylation of mutagenic oligonucleotides, annealing of the phosphorylated oligonucleotides to dU-ssDNA, and enzymatic synthesis of CCC-dsDNA. To phosphorylate mutagenic oligonucleotides, 0.6 μg of each oligonucleotide was mixed with 2 μL of 10X TM buffer (0.1 M $MgCl_2$, 0.5 M Tris, pH 7.5), 2 μL of 10 mM ATP, 1 μL of 100 mM DTT, 2 μL of T4 polynucleotide kinase (10 units/μL, New England Biolabs) and H_2O to a final volume of 20 μL, and incubated for 1 hr at 37 °C. To anneal phosphorylated mutagenic oligonucleotides to dU-ssDNA, 20 μg of dU-ssDNA was mixed with 25 μL of 10X TM buffer, 20 μL of each phosphorylated oligonucleotide, and H_2O to a final volume of 250 μL, and incubated at 90 °C for 3 min, 50 °C for 5 min, and 20 °C for 5 min. To synthesize CCC-DNA, 250 μL of annealed oligonucleotide/template mixture was mixed with 10 μL of 10 mM ATP, 15 μL of 100 mM DTT, 10 μL of dNTP mix (25 mM of each nucleotide), 1 μL of T4 DNA ligase (30 Weiss units/μL, New England Biolabs) and 3 μL of T7 DNA polymerase (10 units/μL, New England Biolabs), and incubated overnight at 20 °C. DNA from the mutagenesis reaction was purified and desalted using a DNA purification kit (Qiagen), and electroporated into *dut*$^{+}$/*ung*$^{+}$ *E. coli* cells (DH10B or SR320). If Kunkel mutagenesis was used for site-directed mutagenesis of phagemids or plasmids, DNA from the mutagenesis reaction was electroporated into DH10B cells, plated as individual colonies, and phagemids/plasmids with desired mutations were isolated and verified using Sanger sequencing. If Kunkel mutagenesis was used for CDR diversification, DNA from the mutagenesis reaction was electroporated into

SR320 cells for phagemid library generation or into M13KO7-infected SR320 cells for library phage production.

4.3 Electroporation of Library DNA and Phage Production

SR320 *E. coli* cells were specially derived for high-efficiency electroporation and large-scale phage production. Electrocompetent SR320 cells and M13KO7-infected electrocompetent SR320 cells were prepared and transformed with library DNA according to established procedures (Fellouse and Sidhu, 2006; Tonikian *et al*., 2007; Nelson and Sidhu, 2012). The following protocol was used to transform M13KO7-infected SR320 cells with library DNA for phage library production. Briefly, 350 µL of highly-concentrated electrocompetent cells (~$3X10^{11}$ CFU/mL) was gently mixed with 50 µL of purified library DNA (20 µg) on ice, transferred to a cold 0.2 cm gap electroporation cuvette, and electroporated using a BTX ECM-600 electroporation system with the following settings: 2.5 kV field strength, 125 Ohms resistance, and 50 µF capacitance. Electroporated cells were recovered immediately with 2 mL of pre-warmed SOC media, transferred into 23 mL of pre-warmed SOC media, and incubated for 30 min at 37 °C with shaking at 200 rpm. The culture was transferred into 500 mL of 2YT media supplemented with 100 µg/mL carbenicillin and 25 µg/mL kanamycin (for phagemid and helper phage selection) and incubated overnight at 37 °C with shaking at 200 rpm. Phage particles were precipitated from the culture supernatant using 1/5 volume of ice-cold PEG/NaCl solution (20% PEG-8000, 2.5 M NaCl), and resuspended in PBS containing 0.5% BSA and 0.05% Tween (PBT). The phage solution was centrifuged again to pellet any insoluble matter. The phage supernatant was quantified using UV spectrometry (OD_{268} = 1.0 for a solution of 5×10^{12} phage/ml) and stored at -80 °C in the presence of protease inhibitors (2%) and sterile glycerol (25%).

4.4 Library-F Amplification

An aliquot of library-F phage was obtained from the Sidhu lab at the University of Toronto. To prepare the starter culture, 250 mL of SR320 *E. coli* cells were grown to mid-log phase (OD_{600} of 0.8) in 2YT media supplemented with 5 µg/mL tetracycline, and incubated with 10^{12} library-F phage for 30 min at 37 °C with shaking at 200 rpm. M13KO7 helper phage was added to the culture to a final concentration of 10^{10} PFU/mL and incubated for another 45 min. The starter culture was then transferred to 8 L of 2YT media supplemented with 100 µg/mL

carbenicillin and 25 μg/mL kanamycin and incubated for 18 hrs at 37 °C with shaking at 200 rpm. Cultures were pelleted, and phages were precipitated and purified from the culture supernatant as described previously. Phage pellets from the final step were resuspended gently in 8 mL sterile PBS and stored at -80 °C in the presence of protease inhibitors (2%) and sterile glycerol (25%).

4.5 Construction of Synthetic Antibody Libraries

4.5.1 Library-S

To create the template phagemid for library-S construction, the pHP153 phagemid encoding the anti-maltose binding protein Fab (Persson *et al*., 2013) was used as the starting phagemid (see Appendix 1 for phagemid map). Desired mutations were incorporated into the Hu4D5-8 Fab framework and CDRs by Kunkel mutagenesis. In the first round of mutagenesis, five oligonucleotides were used to incorporate mutations into the five following regions: CDRL1, CDRL2, CDRH1, CDRH2 and FRM3. In the second round of mutagenesis, two oligonucleotides were used to incorporate NotI restriction sites into CDRL3 and CDRH3. The resulting phagemid was sequence verified and used as the template phagemid for library-S mutagenesis. The variable light and heavy chain sequences of the library-S phagemid are included in Figure 5.7. To introduce CDR diversity, Kunkel mutagenesis was used to repair the NotI sites in the template phagemid and to replace CDRL3 and CDRH3 positions with fixed or degenerate codons encoding the amino acid composition shown in Figure 5.6. Custom-designed CDRL3 and CDRH3 mutagenic oligonucleotides were purchased from Tri-Link Biotechnologies (Appendix 3). CDRL3 was diversified using one oligonucleotide (S-L3-09), whereas CDRH3 was diversified using 20 oligonucleotides, each encoding a different CDRH3 length. Kunkel mutagenesis was performed as described in Section 4.2. Twenty different mutagenesis reactions were required for library-S construction, each reaction representing a CDRH3 length. DNA from the mutagenesis reaction was purified using the PCR cleanup kit (Qiagen) and 10 μg of purified DNA was electroporated into SR320 *E. coli* cells for the synthesis of double-stranded phagemid DNA. The phagemid DNA library was extracted from SR320 cells using the DNA mini-preparation kit (Qiagen). To remove non-mutated template DNA, 20 μg of the phagemid DNA library was digested with NotI (New England Biolabs) according to manufacturer's instructions. DNA from the NotI-digestion reaction was purified using the PCR cleanup kit (Qiagen) and 10

μg of purified library DNA was electroporated into M13KO7-infected SR320 *E. coli* cells for phage production (described in Section 4.3). Phages were purified from the culture supernatant, resuspended in PBS and stored at -80 °C in the presence of protease inhibitors (2%) and sterile glycerol (25%). Phages from 20 sub-libraries were rescued separately and equal number of phages from each sub-library (~5×10^{13} PFU) was mixed together to create the master library.

4.5.2 The Modified-F long-CDR library

The template phagemid for constructing the modified F library was derived from the pHP153 phagemid encoding the anti-maltose binding protein Fab (Persson *et al*., 2013). Kunkel mutagenesis was used to replace CDRL3, CDRH1, CDRH2 and CDRH3 regions of the phagemid with TAA stop codons. The resulting phagemid was sequence verified and used as the template phagemid. The modified-F library was constructed and stored using previously established protocols (Fellouse and Sidhu, 2006; Rajan and Sidhu, 2012). Kunkel mutagenesis was used to simultaneously diversify four CDR regions (L3, H1, H2 and H3) within the template phagemid (diversity design explained in Section 5.2.3). CDRs L3, H1 and H2 were diversified using library-F mutagenic oligonucleotides. CDRH3 was diversified using 17 J_H4-CDRH3 library-F mutagenic oligonucleotides and 10 J_H6-CDRH3 library-S mutagenic oligonucleotides (Appendix 4). Eight different mutagenesis reactions were required for constructing the modified-F library, each reaction representing a set of CDRH3 lengths. DNA from the mutagenesis reaction was purified using the PCR cleanup kit (Qiagen) and 10 μg of purified DNA was electroporated into M13KO7-infected SR320 *E. coli* cells for phage production. Phages were purified from the culture supernatant, resuspended in PBS and stored at -80 °C in the presence of protease inhibitors (2%) and sterile glycerol (25%). Phages from eight sub-libraries were rescued separately and equal number of phages from each sub-library (~5×10^{13} PFU) was mixed together to create the master library.

4.5.3 The EGFR domain II structure-guided Fab library

The pHP153 phagemid encoding the anti-maltose binding protein Fab (Persson *et al*., 2013) was used as the template phagemid for constructing the EGFR domain II structure-guided Fab library. A CDRH3 mutagenic oligonucleotide was designed to possess the EGFR dimerization loop graft (84 bp) with one additional NNC codon flanking the 3′ and 5′ ends of the graft. CDRs L3, H1 and H2 were diversified using library-F mutagenic oligonucleotides

(Appendix 5). Kunkel mutagenesis was used to simultaneously diversify four CDR regions (L3, H1, H2 and H3) within the template phagemid to encode the amino acid composition shown in Figure 5.16. The mutagenesis reaction was electroporated into M13KO7-infected SR320 *E. coli* cells for phage production as described previously (Fellouse and Sidhu, 2006; Rajan and Sidhu, 2012). Phages were purified from the culture supernatant, resuspended in PBS and stored at -80 °C in the presence of protease inhibitors (2%) and sterile glycerol (25%).

4.6 Phage Display Selections

Solid-phase panning of phage-displayed Fab libraries was conducted according to previously established protocols (Fellouse and Sidhu, 2006; Rajan and Sidhu, 2012). Recombinant Fc-tag fused human full-length extracellular domains from Notch-1, Notch-2, Notch-3, Jagged-1, Jagged-2 and EGFR proteins were purchased from R&D Systems and used as selection targets. Phages were precipitated from the frozen master library and resuspended in PBS containing 0.5% BSA and 0.05% Tween (PBT). Target and Fc proteins were immobilized on Maxisorp plates (Nunc) at 5 μg/mL by overnight incubation at 4 °C. Target wells were subsequently blocked with PBS containing 0.5% BSA (PB) for 90 min at RT before washing four times with PBS containing 0.05% Tween (PT). Library phage solution (10^{13} PFU/mL) was added to Fc wells and incubated for 90 min at RT to remove phage with affinity for the Fc-tag. The depleted library was transferred to target wells and incubated for 2 hrs at RT. The plate was washed 8X with PT buffer and phages bound to target wells were eluted with 100 mM HCl. The pH of the eluted phage solution was neutralized with 1 M Tris (pH 8), and used to infect 10 volumes of actively growing XL1-Blue cells in 2YT media supplemented with 5 μg/mL tetracycline. Cultures was incubated for 30 min at 37 °C with shaking at 200 rpm. M13KO7 helper phage was added to the culture to a final concentration of 10^{10} PFU/mL and incubated for another 45 min. Cultures were transferred to 25 volumes of 2YT media supplemented with 100 μg/mL carbenicillin and 25 μg/mL kanamycin and incubated overnight at 37 °C with shaking at 200 rpm. Phages were precipitated and purified from the culture supernatant, and used for subsequent rounds of selection. Also, before the helper phage addition step, 10 μL of phage-infected *E. coli* culture was plated as individual colonies on 2YT-carbenicillin plates for determining the titer of the eluted phage, and for isolation and sequencing of phagemid DNA.

The enrichment in the selection process is measured by comparing the phage titers eluted from the target protein and control proteins.

4.7 Affinity Maturation of Fab DL06

To improve the affinity of Fab DL06, an affinity maturation library was first constructed by soft-randomization of the CDRH3 graft in the DL06 Fab-encoding phagemid using Kunkel mutagenesis. CDRH3 was soft randomized using a mutagenic oligonucleotide in which codons encoding for the dimerization loop (excluding Cys240 and Cys267) were replaced with a mixture comprising 85% of the wild-type base and 5% of each of the other three bases. DNA from the mutagenesis reaction was electroporated into M13KO7-infected SR320 *E. coli* cells for phage production. Phages were purified from the culture supernatant, and used for solid-phase selections against immobilized Fc-tagged EGFR-ECD as described in section 4.6. To obtain Fabs with improved affinity, concentrations of Fc-tagged EGFR-ECD were decreased in successive rounds of selection from 5 to 0.5 μg/mL and single clones were isolated for evaluation by Phage-ELISA. Fab clones deemed EGFR-specific by phage-ELISA (target to control protein binding signal ratio >10) were sequenced and characterized by multi-point ELISA.

4.8 Ion Torrent Sequencing

Ion Torrent sequencing was accomplished by the following steps: PCR amplification of CDR, emulsion PCR on Ion sphere particles (ISPs) and sequencing enriched ISPs on an Ion semiconductor chip (Rothberg *et al*., 2012). To PCR amplify the CDR of interest; we designed primers that hybridize to the fixed framework regions of the phagemid that flank the CDR region. The primers contain barcodes for multiplexing purposes and adapter sequences to facilitate emulsion PCR (Appendix 6). We amplified the CDR of interest using designed primers, checked the purity, concentration and length of PCR products on a 2100 bio-analyzer (Agilent Technologies), prepared the template for emulsion PCR by pooling multiple PCR products, performed emulsion amplification of the amplicon library on the Ion OneTouch 2 instrument (Life Technologies), loaded the enriched ISPs into an Ion Semiconductor chip, and sequenced the loaded ISPs on the V2 Ion Personal Genome Machine (PGM) (Thermo-Scientific) according to manufacturer's instructions.

Ion Torrent sequencing of one diversified CDR was accomplished in three steps. (1) The CDR of interest was PCR amplified from phage selection pools using barcoded Forward and

reverse primers (Appendix 6). The PCR reaction mix (50 μL) contained 32.5 μL of nuclease-free H_2O, 10 μL of 5X Phusion High-Fidelity buffer (New England BioLabs), 1 μL of dNTP mix (10 mM of each nucleotide), 1 μL of phage solution (10^{12} PFU/mL), 2.5 μL of 10 μM Forward primer, 2.5 μL of 10 μM Reverse primer and 0.5 μL of Phusion Hot-Start Flex DNA Polymerase (New England BioLabs). The reaction mix was subjected to PCR using the following conditions: initial denaturation at 98 °C for 30 sec, 25 amplification cycles each consisting of a denaturing step at 98 °C for 10 sec, an annealing step at 56 °C for 10 sec, and an extension step at 72 °C for 5 sec, and a final extension at 72 °C for 15 sec. (2) PCR amplicons were purified, quantified, multiplexed, and subjected to emulsion PCR using the Ion OneTouch template kit. (3) Enriched ISPs were loaded on an Ion 314 chip and sequenced using the Ion PGM supplies kit.

Ion Torrent sequencing of the L3-H3 CDR strip was accomplished in six steps. (1) ssDNA was extracted from amplified phage selection outputs (10^{13} PFU) using the Spin M13 kit. (2) 500 ng of ssDNA was subjected to Kunkel mutagenesis for deleting the framework regions between diversified CDRs. In the mutagenesis reaction, one oligonucleotide, L3-H3 Seq (Appendix 6), was used to link the L3-H3 regions together. Phosphorylation of L3-H3 Seq, annealing of L3-H3 Seq to the ssDNA template, and *in vitro* synthesis of CCC-dsDNA were carried out as described in section 4.2. (3) DNA from the mutagenesis reaction was run on an agarose gel and the right-sized product (CCC-dsDNA) was excised and purified using a gel-extraction kit (Qiagen). (4) The L3-H3 CDR strip was PCR amplified from the purified CCC-dsDNA template using barcoded L3-Fwd and H3-Rev primers (Appendix 6). The PCR reaction mix (50 μL) contained 28.5 μL of nuclease-free H_2O, 10 μL of 5X Phusion High-Fidelity buffer, 1 μL of dNTP mix, 5 μL of CCC-dsDNA (50 ng), 2.5 μL of 10 μM L3-Fwd, 2.5 μL of 10 μM H3-Rev, and 0.5 μL of Phusion Hot-Start Flex DNA Polymerase. The reaction mix was subjected to PCR using the following conditions: initial denaturation at 98 °C for 30 sec, 25 amplification cycles each consisting of a denaturing step at 98 °C for 10 sec, an annealing step at 56 °C for 10 sec, and an extension step at 72 °C for 5 sec, and a final extension at 72 °C for 15 sec. (5) PCR amplicons were purified, quantified, multiplexed and subjected to emulsion PCR using the Ion PGM Template OT2 200 kit. (6) Enriched ISPs were loaded on an Ion 314 Chip and sequenced using the Ion PGM Sequencing 200 V2 kit.

Ion Torrent sequencing of the L3-H1-H2-H3 CDR strip was accomplished in six steps. (1) ssDNA was extracted from amplified phage selection outputs (10^{13} PFU) using the Spin M13

kit. (2) 500 ng of ssDNA was subjected to Kunkel mutagenesis for deleting the framework regions between four diversified CDRs. In the mutagenesis reaction, three oligonucleotides (L3-H1 Seq, H1-H2 Seq, H2-H3 Seq, were used to link the L3-H1-H2-H3 regions together (oligonucleotide sequences in Appendix 6). Phosphorylation of oligonucleotides, annealing of oligonucleotides to the ssDNA template, and *in vitro* synthesis of CCC-dsDNA were carried out as described in section 4.2. (3) DNA from the mutagenesis reaction was run on an agarose gel and the right-sized product (CCC-dsDNA) was excised and purified using a gel-extraction kit (Qiagen). (4) The L3-H1-H2-H3 CDR strip was PCR amplified from 50 ng of purified CCC-dsDNA using barcoded L3-Fwd and H3-Rev primers (see reaction setup and conditions above). (5) PCR amplicons were purified, quantified, multiplexed and subjected to emulsion PCR using the Ion PGM Template OT2 400 kit. (6) Enriched ISPs were loaded on an Ion 314 Chip and sequenced using the Ion PGM Sequencing 400 kit.

4. 9 NGS Data Processing and Analysis

Sequencing flowgrams were base-called, and subsequently adapter regions were removed in the Ion Torrent server. Sequences were parsed by barcodes in the Geneious software (Kearse *et al*., 2012) and the following actions were performed on the resulting FASTQ files in the Galaxy server (Goecks *et al*., 2010; Blankenberg *et al*., 2010): (1) FASTQ quality trimmer (trims from 3' end with a window size of 1 and step size of 1, trimming continues until the base has a quality score of 17 or greater); (2) FASTQ quality filter (filters and removes sequences that are smaller than the smallest possible CDR or have a combined quality score less than 17); (3) FASTQ to FASTA (removes the quality information from the file to allow analysis with downstream tools); (4) CDR clip filter (searches for a string of bases following the CDR of interest, sequences that contain the clipping sequence are retained, the clipping sequence and the three bases that follow it are retained, sequences that do not contain the clipping sequence are removed); (5) Translate (sequences are translated to amino acids); (6) FASTA to Tabular (file format is converted to allow analysis with downstream tools); (7) Frame-shift filter (searches for sequences that contain the four specified amino acids that follow the CDR of interest, sequences that do not contain these amino acids are removed); (8) Tabular to FASTA (file format is converted to allow analysis with downstream tools); (9) Reverse translate (reverse translates amino acids to nucleotides for downstream analysis); (10) Trim CDR clips (searches for and trims the sequences that immediately precede and follow the CDR of interest, sequences that do

not contain clips are removed); (11) Count sequences (this tool searches for and counts the number of template CDR sequences, sequences with stop codons, unique CDRs and determines their relative frequencies); and (12) Output Tabular and FASTA files (the Galaxy server outputs Tabular and FASTA files for further analysis). Further NGS analyses were conducted using Excel (Microsoft), Geneious (Kearse *et al*., 2012), WebLogo (Crooks *et al*., 2004), and R (R Core Team, 2013) platforms.

4.10 Expression and Purification of Fabs

Fab sequences were sub-cloned from the phagemid vector into a modified pCW-LIC Fab expression vector (plasmid map in Appendix 2) using standard molecular biology procedures. Briefly, Fab sequences were amplified from phagemids by PCR, and ligated into the SacI/XhoI-digested pCW-LIC vector using Gibson assembly (Gibson *et al*., 2009). The Fab expression plasmid was sequence verified and electroporated into BL21 *E. coli* cells. Transformed *E. coli* cells were grown to mid-log phase in 2YT media supplemented with 100 μg/mL carbenicillin. The starter culture (4 mL) was transferred into 400 mL of Overnight Express Terrific Broth (TB) auto-induction medium (EMD Millipore) supplemented with 100 μg/mL carbenicillin and incubated for 18 hrs at 25 °C with shaking at 200 rpm. Cells were pelleted by centrifugation, and lysed in protein-L binding buffer (20 mM Na_2HPO_4, 0.15 M NaCl, pH 8) containing 1:100 dilution of protease inhibitor cocktail (sigma) using a cell disruptor (Constant Systems). Clarified supernatant was loaded into a Protein-L column (GE healthcare) and washed with 10 column volumes of Protein-L binding buffer (AKTA prime plus). Fabs were eluted with IgG elution buffer (Thermo-Scientific) and neutralized with 1 M Tris-HCl (pH 9). Eluted Fabs were dialyzed against PBS and stored at -20 °C. Fab purity was verified using 2100 bio-analyzer (Agilent Technologies) and Fab concentration was determined by UV-visible spectrometry.

For high-throughput Fab expression and purification, Fab sequences were sub-cloned from the phagemid vector into the pCW-LIC Fab expression vector in 96-well format. To eliminate the use of Sanger sequencing, Gibson assembly reactions were directly electroporated into BL21 *E. coli* cells, and three colonies from each reaction were screened for Fab expression using bio-layer interferometry in 96-well format. Briefly, single colonies were transferred to 1 mL of TB auto-induction medium supplemented with 100 μg/mL carbenicillin in 96-well deep well boxes, and incubated for 18 hrs at 25 °C and 200 rpm. Cells were pelleted by centrifugation

and lysed with 200 μL of B-PER bacterial protein extraction reagent (Pierce). Cells were centrifuged again, and 50 μL of the clarified supernatant was transferred to 384-well plates. Fab expression was detected using anti-Fab C_H1 biosensors and anti-HIS biosensors in the Octet RED384 system (ForteBio) according to manufacturer's instructions. Positive clones were transferred into 30 mL of TB auto-induction medium supplemented with 100 μg/mL carbenicillin, and incubated for 18 hrs at 25°C with shaking at 200 rpm. Cells were pelleted by centrifugation, and lysed in protein-L binding buffer supplemented with protease inhibitors using a cell disruptor. Clarified supernatant was incubated with 200 μL of Protein-L resin (GenScript) for 1 hr at 4 °C on a nutator. The Protein-L resin was collected by centrifugation and washed 5X with Protein-L binding buffer. Fabs were eluted with IgG elution buffer and neutralized with 1 M Tris-HCl (pH 9). Eluted Fabs were dialyzed against PBS and stored at -20 °C. Fab purity was verified using 2100 bio-analyzer and Fab concentration was determined by UV-visible spectrometry.

For purification of biotinylated Fabs, Fab sequences were sub-cloned into RH2.2 Fab expression vector using standard restriction enzyme cloning procedures. The vector contains an Avi tag (GGGLNDIFEAQKIEWHE) at the C-terminus of the Fab heavy chain. Expression constructs were sequence verified and electroporated into BL21 *E. coli* cells previously transformed with a biotin ligase-encoding plasmid. Single clones were grown to mid-log phase in 2YT media supplemented with 100 μg/mL carbenicillin and 5 μg/mL chloramphenicol. Fabs were expressed in TB auto-induction media supplemented with the same selection antibiotics plus 25 mM biotin. Fabs were purified, quantified and stored as described previously.

4.11 Enzyme-linked immunosorbent assays

Phage-ELISA was performed to check the binding of phage-displayed Fabs to immobilized target proteins. The phagemid-encoding Fab was electroporated into M13KO7-infected electrocompetent SR320 *E. coli* cells for phage production. The cells were rescued with pre-warmed SOC media and incubated for 30 min at 37 °C. The culture was transferred to 30 mL of 2YT media supplemented with 100 μg/mL carbenicillin and 25 μg/mL kanamycin and incubated overnight at 37 °C with shaking at 200 rpm. Phages were precipitated from the culture supernatant using 6 ml of ice-cold PEG/NaCl solution, resuspended in PBT buffer, and quantified using UV spectrometry.

To conduct phage-ELISA, target and control proteins were immobilized at 5 μg/mL on Maxisorp plates (Nunc) by overnight incubation at 4 °C. The wells were subsequently blocked with PB buffer for 90 min at RT before washing four times with PT buffer. Wells were then exposed to PBT-diluted phage solution (10^{12} PFU/mL) for 30 min, washed 8X with PT buffer, and then incubated with a 1:3000 dilution of HRP-conjugated anti-M13 antibody (GE healthcare) for 30 min at RT. Plates were washed again 6X with PT buffer and 2X with PBS. Wells were developed with 3,3′,5,5′-tetramethylbenzidine (TMB) substrate for 5 min, and quenched with equal volume of 1 M H_3PO_4. The plates were read at 450 nm using a SpectraMax 340PC plate reader (Molecular devices).

Fab-ELISA was performed to check the binding of purified Fabs to immobilized target proteins. Proteins were immobilized at 5 μg/mL on Maxisorp plates (Nunc) by overnight incubation at 4 °C. The wells were subsequently blocked with PB buffer for 90 min at RT before washing 4X with PT buffer. Wells were then exposed to 100 μL of Fab solution diluted in PT for 30 min, washed 10X with PT buffer, and then incubated with a 1:3000 dilution of HRP-conjugated anti-His antibody (Rockland Biosciences) for 30 min at RT. Plates were washed again 6X with PT buffer and 2X with PBS. Wells were developed with TMB substrate for 5 min, and quenched with equal volume of 1 M H_3PO_4. The plates were read at 450 nm using a SpectraMax 340PC plate reader (Molecular devices). Single-point Fab-ELISA was used to assess Fab specificity, and multi-point Fab-ELISA was used to calculate the EC_{50} for Fab binding to the immobilized target. In multi-point Fab ELISA, ABS_{450} values were obtained for a range of Fab concentrations and EC_{50} was calculated by fitting the data to the one-site specific-binding equation in Prism (Graphpad).

4.12 Analysis of Fab Binding Kinetics

The ForteBio Octet RED384 system (Pall Corporation) was used to measure the binding kinetics between purified Fabs and target proteins. Fabs were immobilized on amine-reactive generation-2 biosensors (for K_D <5 nM) or anti-Fab CH1 biosensors (for K_D >5 nM) according to manufacturer's instructions. Immobilized Fabs were exposed to increasing concentrations of target proteins, and association and dissociation rates were measured by the shift in wavelength (nm). All reactions were performed at 25 °C in PBS. For each sensor-immobilized Fab, at least four different target protein concentrations were used, and K_D (equilibrium dissociation constant)

was obtained by fitting the data to 1:1 binding model. Data processing and curve fitting were performed using the Octet Software (ForteBio).

4.13 Characterization of Anti-EGFR Fabs and Competition Assays

Fab-ELISA was used to test the specificity of anti-EGFR Fabs (DL06 and H) against the ErbB family receptors. Recombinant Fc-tagged ectodomains of ErbB proteins were immobilized on Maxisorp plates at 5 μg/mL by overnight incubation at 4 °C. Wells were blocked, washed and incubated with anti-EGFR Fabs (100 nM) for 30 min at RT. Plates were washed and incubated with 400 nM of HRP-conjugated anti-FLAG secondary antibody (Sigma) for 30 min at RT. Plates were washed, developed and read at 450 nm as described previously.

To measure relative contributions of diversified CDRs in Fab-DL06 towards EGFR binding, we back-mutated sequences of each diversified CDR in Fab-DL06 to the corresponding anti-MBP Fab CDR and evaluated the binding of these Fabs to EGFR-ECD by Fab-ELISA. First, each diversified CDR in the DL06-encoding RH2.2 Fab expression vector was mutated to anti-MBP CDR using Kunkel mutagenesis. CDR-mutated Fab-DL06 expression plasmids were sequence-verified and electroporated into BL21 *E. coli* cells. CDR-mutated DL06 Fabs were expressed and purified as described previously. To conduct Fab-ELISA, EGFR-ECD was immobilized on Maxisorp plates by overnight incubation at 4 °C. Wells were blocked, washed and incubated with WT and CDR-mutated DL06 Fabs (250nM) for 30 min at RT. Plates were washed and incubated with 400 nM of HRP-conjugated anti-FLAG secondary antibody (Sigma) for 30 min at RT. Plates were washed, developed and read at 450 nm as described previously.

Competitive-ELISA was used to test whether Fab-DL06 and Fab-H compete with each other for binding EGFR-ECD. The assay was performed using purified Fabs in both non-biotinylated and site-specifically biotinylated formats (BT-Fab). EGFR-ECD was immobilized on Maxisorp plates by overnight incubation at 4 °C. Wells were blocked, washed and incubated with a saturating concentration of Fab-DL06 (50 μg/mL) for 30 min, followed by direct addition of BT-Fab-H at a sub-saturating concentration (100 ng/mL). The assay was also conducted in the opposite configuration in which EGFR-ECD-coated wells were incubated with a saturating Fab-H concentration (50 μg/mL) for 30 min, followed by direct addition of BT-Fab-DL06 at a sub-saturating concentration (250 ng/mL). Plates were incubated for 30 min, washed, and binding of BT-Fab to EGFR-ECD was detected using 0.1 μg/mL of HRP-conjugated streptavidin prepared in PBT. Plates were washed, developed and read at 450 nm as described previously.

Competitive-ELISA was also used to test whether anti-EGFR mAbs block the binding of Fab-DL06 and Fab-H to EGFR-ECD. EGFR-ECD was immobilized on Maxisorp plates, blocked, washed and incubated with anti-EGFR antibodies at a saturating concentration (65 nM) for 30 min at RT. Anti-EGFR Fabs were added to the wells containing the EGFR-antibody complex and incubated for 30 min at RT. Plates were washed and incubated with 400 nM of HRP-conjugated anti-FLAG secondary antibody (Sigma) for 30 min at RT. Plates were washed, developed and read at 450 nm as described previously.

Phage-ELISA was used to test whether phage-displayed EGFR domain III binds to anti-EGFR Fabs and antibodies (Tundidor *et al*., 2014). To display EGFR domain III on phages, we sub-cloned EGFR domain-3 (residues 311-514) into the pHP153 phagemid using Gibson assembly (Gibson *et al*., 2009). Following sequence verification, the EGFR domain III phagemid was electroporated into M13KO7-infected SR320 *E. coli* cells for phage production. Phage was precipitated, purified and quantified as described previously. Fabs, antibodies and control proteins were immobilized on Maxisorp plates at 5 μg/mL by overnight incubation at 4 °C. The wells were blocked, washed and incubated with PBT-diluted EGFR domain III phage solution (10^{12} PFU/mL) for 30 min. The wells were washed and incubated with a 1:3000 dilution of HRP-conjugated anti-M13 antibody for 30 min at RT. Plates were washed, developed and read at 450 nm as described previously.

4.14 Flow Cytometry and Epitope Mapping Studies

Flow cytometry was used: (1) to obtain EC_{50} values for anti-EGFR Fabs and mAbs binding to cell-surface EGFR; (2) to test the influence of EGF and anti-EGFR mAbs on the binding of anti-EGFR Fabs to cell-surface EGFR; and (3) to test the binding of anti-EGFR Fabs and mAbs to cell-surface EGFR mutants and truncations. Two cell lines were used in this work: HEK293F cells transiently transfected with EGFR and A431 cells that endogenously express EGFR. HEK293F cells were maintained in suspension in Freestyle 293 expression media at 37 °C and 5% CO_2 with shaking at 125 rpm. A431 cells were maintained as monolayer cultures in 90% Eagle's minimum essential media with Earle's balanced salt solution (EMEM/EBSS) and 10% fetal bovine serum (FBS) at 37 °C and 5% CO_2. Cells were cultured until 80-90% confluent, and passaged every 3-4 days. For overexpressing EGFR in HEK293F cells, we used an EGFR plasmid (Addgene) that encodes for full-length human EGFR fused to green fluorescent protein (GFP) at its C-terminus (Carter and Sorkin, 1998).

To transfect HEK293F cells, 1.5 X 10^6 cells were transferred from flasks to 10 cm plates containing 10 mL Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS, and maintained at 37 °C and 5% CO_2. On the day prior to transfection, cells were transferred to 10 mL serum-free DMEM and maintained under the same conditions. For transfection, 10μg of plasmid DNA was added to a mixture of 36 μL polyethylenimine (1mg/mL pH 7.5) and 540 μL warm optimal-minimal essential medium, and incubated at RT for 30 minutes, before drop-wise addition to plates containing HEK293F cells. Plates were incubated for 24 hrs at 37 °C and 5% CO_2. Media was changed to DMEM with 10% FBS and expression was allowed to proceed for another 24 hrs.

To obtain EC_{50} values for anti-EGFR Fabs binding to cell-surface EGFR, EGFR-expressing HEK293F or A431 cells were split into equal fractions (~2X10^5 cells) and incubated with different concentrations of 100 μL Fab-DL06 or Fab-H prepared in PBS for 1 hr at 4 °C. Cells were washed with PBS, and incubated with 100 μL of PE-conjugated anti-FLAG mouse secondary antibody (1:5000 dilution, Prozyme) in the dark on ice for 30 min with occasional mixing. Flow cytometric analysis of binding was conducted on a MACSQuant VYB flow cytometer (Miltenyi-Biotec). A minimum of 10^3 cell events was acquired for each Fab concentration. GFP fluorescence was used to quantify EGFR expression and PE fluorescence was used to quantify Fab binding to cells. PE-labeled cells were compared to un-stained control cells or cells labeled only with the secondary antibody. The proportion of PE-positive cells was determined from the subpopulation of high-GFP expressing cells (~50%) using the FlowJo Software. To obtain binding curves and affinity estimates, log [Fab] was plotted vs. %PE-positive cells and fit using the log [inhibitor] vs. response equation with standard fit in Prism (Graphpad).

To obtain EC_{50} values for anti-EGFR mAbs binding to cell-surface EGFR, EGFR-expressing HEK293F cells were split into equal fractions (~2X10^5 cells) and incubated with different concentrations of 100 μL anti-EGFR mAb prepared in PBS for 1 hr at 4 °C. Cells were washed with PBS, stained with 100 μL of PE-conjugated goat anti-human IgG secondary antibody (1:400 dilution, Beckman Coulter), and analyzed by flow cytometry as described previously.

To test the influence of EGF on anti-EGFR Fabs binding to cell-surface EGFR, EGFR-expressing HEK293F cells were harvested, incubated with 10 nM EGF for 15 min at 4 °C, split

into aliquots of ~5 X 10^5 cells, and incubated with serial dilutions of either Fab-DL06 or Fab-H for 1 hr on ice with occasional gentle swirling. Cells were washed with PBS, stained with the PE-conjugated anti-FLAG secondary antibody, and analyzed by flow cytometry as described previously.

To test the influence of anti-EGFR mAbs (CTX and PMB) on Fab-DL06 and Fab-H binding to cell-surface EGFR, EGFR-expressing HEK293F cells were harvested, incubated with 5 nM CTX or PMB for 15 min at 4 °C, split into aliquots of ~5 X 10^5 cells, and incubated with serial dilutions of either Fab-DL06 or Fab-H for 1 hr on ice with occasional gentle swirling. Cells were washed with PBS, stained with the PE-conjugated anti-FLAG secondary antibody, and analyzed by flow cytometry as described previously.

To conduct epitope mapping of anti-EGFR Fabs, we generated EGFR mutants by Kunkel mutagenesis, and tested the binding of anti-EGFR Fabs to mutant EGFR constructs by flow cytometry. The EGFR-GFP plasmid encodes full-length human EGFR comprising the extracellular domains (residues 1-622), the transmembrane domain (residues 622-644) and the intracellular kinase domain (residues 644-697). WT-EGFR was expressed using this construct without any modifications. The presence of f1 origin in the plasmid facilitated the isolation of template dU-ssDNA using CJ236 *E. coli* cells. Mutations were introduced via annealing, extension and ligation of phosphorylated mutagenic oligonucleotides on the EGFR-GFP dU-ssDNA template using Kunkel mutagenesis as described previously (Tonikian *et al*., 2007; Nelson and Sidhu, 2012). The following EGFR mutants were generated: an EGFR double mutant (Y251A + R285S), the EGFR dimerization loop deletion mutant (Δ240-267), and the EGFR vIII mutant (Δ6-274). The following single and multi-domain constructs were also generated: D1 (residues 1-162), D2 (residues 163-311), D3 (residues 312-480), D4 (residues 481-614), D1-D2 (residues 1-311), D2-D3 (residues 163-480), D3-D4 (residues 312-614), D1-D2-D3 (residues 1-480) and D2-D3-D4 (residues 163-614). EGFR residues 615-1186 (comprising the extracellular linker- transmembrane- intracellular linker- kinase domain- C-terminus tail) and the GFP region were not manipulated. All new EGFR constructs were sequence verified, amplified in DH10B *E. coli* cells, and used for transfecting HEK293F cells. To test the binding of anti-EGFR Fabs to EGFR mutants, cells were split into equal fractions, incubated with different concentrations of 100 µL anti-EGFR Fab, washed with PBS, stained with the PE-conjugated anti-FLAG secondary antibody, and analyzed by flow cytometry. To test the binding of anti-EGFR mAbs to EGFR

mutants, cells were split into equal fractions, incubated with different concentrations of 100 μL anti-EGFR mAb, washed with PBS, stained with the PE-conjugated goat anti-human IgG secondary antibody, and analyzed by flow cytometry.

4.15 EGFR Signaling Assays

Western analysis was used to evaluate the effects of anti-EGFR Fabs on EGF-induced receptor phosphorylation in A431 cells and MDA-MB-231 cells. A431 cells were cultivated until 80-90% confluent in 90% RPMI medium supplemented with 10% FBS at 37 °C and 5% CO_2. MDA-MB-231 cells were cultivated until 80-90% confluent in 90% DMEM medium supplemented with 10% FBS at 37 °C and 5% CO_2. Growth medium was replaced with pre-warmed serum-free medium and incubated for 24 hrs. Serum-starved A431 and MDA-MB-231 cells were pre-treated with various concentrations of Fabs in serum-free media and incubated at 37 °C for 1 hr. Cells were stimulated with 50 ng/mL EGF for 15 min in the presence of Fabs. Cells were washed 2X with cold PBS and collected by scraping with 200 μL of lysis buffer and 10 μL of protease/phosphatase inhibitors. Cells were sonicated 2X and cell debris was removed by centrifugation. A 20 μL aliquot of cleared lysate was mixed with 4X loading buffer, separated by electrophoresis on a 4-20% SDS-PAGE gel, and transferred to a PVDF membrane using standard methods. Blots were blocked with 5% BSA overnight, and incubated with 1:1000 dilution of rabbit polyclonal anti-phospho-EGFR Tyr1173 antibody (Cell Signaling) or rabbit polyclonal anti-EGFR antibody (Cell Signaling) for 24 hrs. Blots were washed 3X with PT buffer, and incubated with a 1:5000 dilution of anti-rabbit-HRP secondary antibody (Santa Cruz) for 30 min. Blots were developed with a chemiluminescent substrate (Thermo Scientific) and exposed to CL-XPosure radiography film (Thermo Scientific) for 1 min. Intensity of bands were quantified using the ImageLab software (Bio-Rad).

5. RESULTS AND DISCUSSION

This chapter is divided into three sections. **Section 5.1** is entitled "A platform for high-throughput reconstruction of synthetic antibody fragments from phage selection outputs". In this section, we describe the design, construction and validation a new single-framework synthetic Fab library, named library-S. We developed a method for linking and sequencing all diversified CDRs in phage Fab pools by Ion Torrent sequencing. We showed that low-frequency rare clones reconstructed from NGS information could possess higher affinity and better specificity than high-frequency top clones isolated by Sanger sequencing. **Section 5.2** is entitled "Systematic generation of synthetic antibody fragments with optimal CDR lengths for Notch-1 recognition". In this section, we tested the potential of two single-framework synthetic Fab libraries to generate Fabs with high affinity and specificity for Notch-1. Over the course of Fab generation, we also showed that implementing NGS approaches, screening focused diversity libraries, fine-tuning the library diversity and making small changes to diversity designs can improve the success rate of single-framework synthetic Fab libraries. **Section 5.3** is entitled "Generation and validation of anti-EGFR Fabs from naïve and structure-guided synthetic antibody libraries". In this section, we describe the design and construction of a structure-guided Fab library that was biased towards interacting with EGFR domain-II. We isolated two anti-EGFR Fabs, one from the structure-guided synthetic Fab library, and one from a previously described naïve synthetic Fab library. We characterized both anti-EGFR Fabs in terms of affinity, specificity, activity and binding site location. The motif-grafting project described in section 5.3 is a collaboration between the Geyer lab at the University of Saskatchewan and the Sidhu lab at the University of Toronto. Contributions of other researchers towards this PhD thesis are included in Appendix 7.

5.1 A Platform for High-Throughput Reconstruction of Synthetic Antibody Fragments from Phage Selection Outputs

The objective of this section is to develop an NGS-assisted antibody discovery platform by integrating phage-displayed single-framework synthetic Fab libraries with Ion Torrent sequencing. To accomplish this, we first developed a method for linking and sequencing all diversified CDRs in phage Fab pools without losing the CDR pairing information. Second, we designed and constructed a new phage-displayed single-framework synthetic Fab library, named library-S. Third, we analyzed library-S quality by Ion Torrent sequencing and validated its functionality by generating high-affinity Fabs against Jagged-1 and Jagged-2. Fourth, we validated the NGS-assisted Fab reconstruction platform by reconstructing and testing low-frequency rare clones from phage selection outputs. While previous studies chose rare clones for rescue based on their relative frequencies in sequencing outputs, we chose rare clones for reconstruction from less-frequent CDRH3 lengths. In some cases, reconstructed rare clones (frequency ~0.1%) showed higher affinity and better specificity than high-frequency top clones isolated by Sanger sequencing.

5.1.1 Integrating Antibody Phage Display with Ion Torrent Sequencing

Ion Torrent sequencing consists of three basic steps: (1) PCR amplification of a short region of interest, (2) emulsion PCR on proprietary Ion sphere particles (ISPs); and (3) sequencing enriched ISPs on an Ion semiconductor chip (Rothberg *et al*., 2012). Ion Torrent chips offer three different read lengths (100, 200 or 400 bp). To sequence a CDR of interest, we used 100 bp chips. For example, to amplify CDRH3 from phage pools, we designed primers that hybridize to the fixed framework regions of the phagemid that flank the CDRH3 region. We introduced barcodes into the forward primer for multiplexing different samples in one run. The forward and reverse primers also contained adapter sequences to facilitate emulsion PCR. A truncated P1 (trP1) adapter in the reverse primer hybridizes to a complimentary trP1 sequence on ISPs, and an A-adapter in the forward primer contains a starting sequence for the synthesis of complimentary DNA during emulsion PCR. We amplified the CDR of interest from selection outputs using designed primers, performed emulsion amplification of the amplicon library on Ion sphere particles (ISPs), and sequenced the enriched ISPs on a 100 bp Ion semi-conductor chip. The strategy for sequencing a CDR of interest is illustrated in **Figure 5.1**. We built a custom

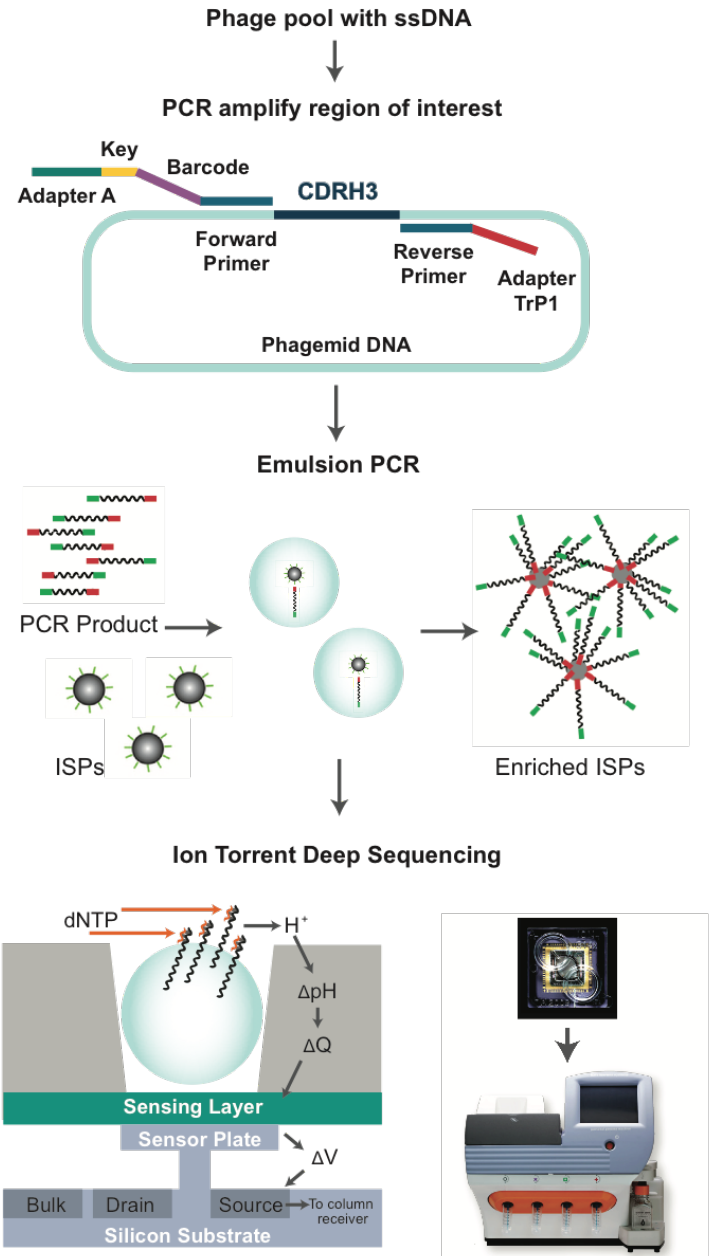


Figure 5.1: Workflow for CDRH3 Ion Torrent sequencing. The CDRH3 region is PCR amplified from phage pools using primers containing barcodes (for multiplexing) and adaptors (for emulsion PCR). Amplified DNA is subjected to emulsion PCR for clonal amplification of individual DNA molecules on Ion sphere particles (ISPs). Following emulsion PCR, each ISP contains up to a million copies of a one short DNA of interest. Enriched ISPs are loaded into an Ion semiconductor chip that contains millions of micro-wells. Each well takes up an enriched ISP and acts as a pH sensor. The chip is placed on the Ion personal genome machine (Ion PGM), which sequentially floods the chip with one nucleotide after another. When a nucleotide complements the DNA sequence in a particular well, DNA polymerase incorporates it to the growing strand. This results in the release of hydrogen ions and a change in pH of the well. An ion sensor detects this pH change and translates the chemical signal into a digital signal.

workflow for NGS data processing and analysis (see Section 4.9). In brief, base-called reads were parsed by barcodes, quality trimmed and filtered. Filtered sequences were translated, identified, counted and analyzed in terms of length and amino acid distribution.

Our libraries contain up to four diversified CDRs in a single fixed Fab-4D5 framework (Fellouse *et al*., 2007). The four diversified CDRs are only 200 bp in length combined together, which is shorter than the read length offered by the 400 bp Ion semiconductor chip. Therefore, we sought to link the diversified CDRs in phage pools next to each other by deleting the intervening framework regions, and sequence all four diversified CDRs as a CDR strip (L3-H1-H2-H3) on a 400 bp chip. Since the template for framework deletion is phage-derived single-stranded phagemid DNA (ssDNA), we chose to use Kunkel mutagenesis; a method that introduces site-directed mutations when ssDNA is converted into double-stranded DNA *in vitro* (Kunkel *et al*., 1987). The overall strategy for CDR strip generation and sequencing is illustrated in **Figure 5.2**. We generated two kinds of CDR strips using our method, L3-H3 and L3-H1-H2-H3. Originally, we developed the method to generate L3-H3 strips from a library containing only two diversified CDRs (library-S), and to sequence them on a 200 bp chip. As the Ion Torrent sequencing technology improved, we extended our method to generate L3-H1-H2-H3 strips from a library containing four diversified CDRs (library-F), and to sequence them on a 400 bp chip. The oligonucleotides used in the Kunkel reaction are 30 bases long and anneal to the 3' region of one CDR and the 5' region of the adjacent CDR within the same ssDNA. For example, out of 30 bases of the L3-H1 primer, 15 bases anneal to the 3' region of CDRL3 and 15 bases anneal to the 5' region of CDRH1. One oligonucleotide was used to bring the L3-H3 regions together, and three oligonucleotides were used to bring the L3-H1-H2-H3 regions together. The mutagenesis reaction was run on an agarose gel, and the desired product was excised. The gel-purified product was then used as a template for PCR amplification of the CDR strip. Quantified, multiplexed amplicons were subjected to emulsion PCR and Ion Torrent sequencing.

To reconstruct Fab clones from CDR strip sequencing information, we cloned desired CDR combinations into the Fab-4D5 encoding phagemid by Kunkel mutagenesis **(Figure 5.3)**. A Fab-4D5 phagemid whose CDRs have been replaced with NotI sites was used as a template to reconstruct Fab clones. Primers were designed to encode for a desired CDR sequence and to hybridize to either side of the CDR. Two or four primers were used to reconstruct Fab clones from library-S or library-F selections, respectively. Following mutagenesis, the reaction was

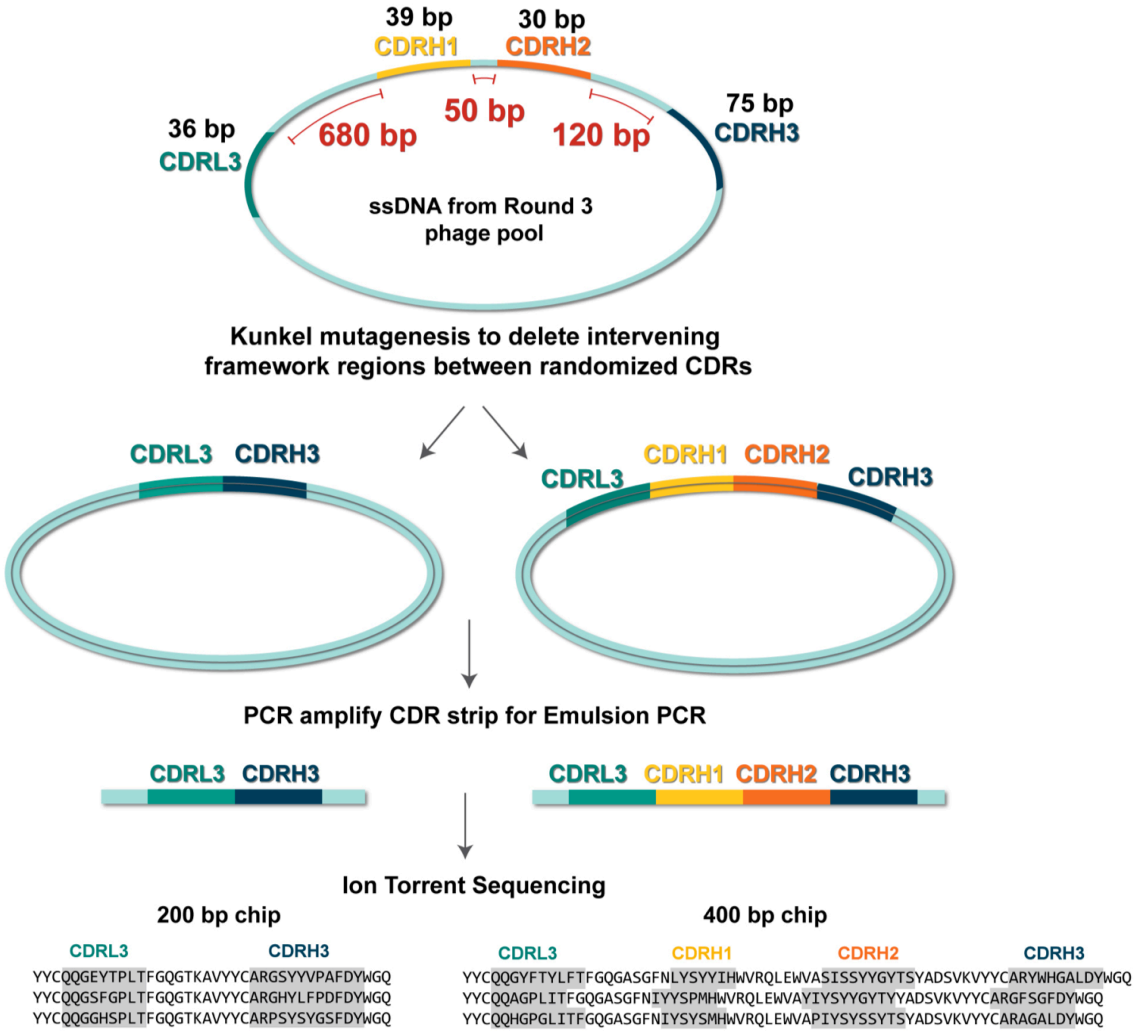


Figure 5.2: Strategy for CDR strip generation and sequencing. Single-stranded DNA (ssDNA) rescued from round-3 phage pools is subjected to Kunkel mutagenesis for deleting the intervening framework regions between diversified CDRs. This step links the diversified CDRs next to each other (L3-H3 from library-S selections and L3-H1-H2-H3 from library-F selections) without losing the CDR pairing information. The right-sized product from the mutagenesis reaction is used as a template for PCR amplification of the CDR strip. Quantified, multiplexed amplicons are subjected to emulsion PCR and Ion Torrent sequencing. L3-H3 and L3-H1-H2-H3 strips are sequenced on 200 bp and 400 bp chips, respectively.

transformed into *E. coli* and positive clones were screened by NotI digestion. It is worth noting the differences between Kunkel reactions used in Figures 5.2 and 5.3. The first Kunkel reaction was used to link the diversified CDRs together with fixed primer sequences (ssDNA template was variable), and the second Kunkel reaction was used to reconstruct Fab phagemids using variable primers (ssDNA template was constant for all Fabs). The second one was a very typical

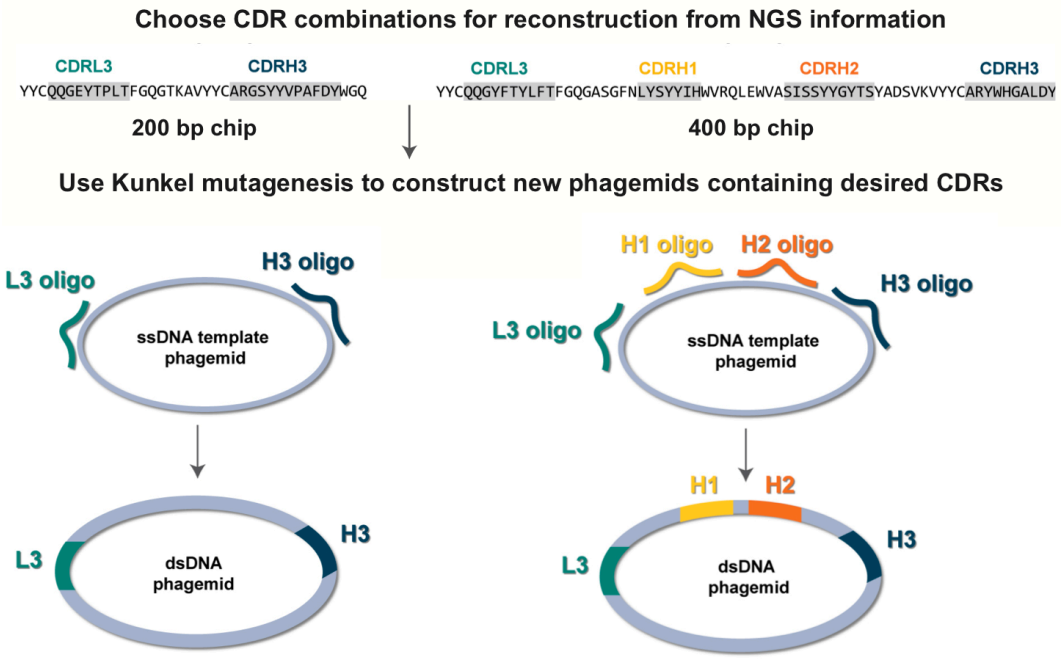


Figure 5.3: Strategy for reconstructing Fab clones from NGS information. CDR strips (L3-H3 or L3-H1-H2-H3) generated from round-3 phage pools are subjected to Ion Torrent Sequencing. Following NGS analysis, desired CDR combinations are reconstructed by cloning CDR-encoding oligonucleotides into the Hu4D5 Fab-encoding template phagemid by Kunkel mutagenesis.

Kunkel reaction; uracil-containing ssDNA was used as a template, and the mutagenesis reaction was transformed into *E. coli* to eliminate the undesired wild-type strand. Since ssDNA for the first reaction came from phage selection pools, it was not uracil-inserted, and the desired product (CCC-dsDNA) from the mutagenesis reaction was isolated by gel purification.

5.1.2 Design of Synthetic Antibody Libraries F and S

We used two single-framework synthetic Fab libraries (F and S) for validating the NGS-assisted Fab reconstruction platform. Library-F is a highly validated Fab library that has been used successfully against diverse kinds of targets (Persson *et al*., 2013; Hornsby *et al*., 2015; Zhong *et al*., 2015 and Na *et al*, 2016). It was built on the Hu4D5-8 Fab scaffold derived from the anti-Her2 antibody Herceptin. The Hu4D5-8 Fab framework is very stable (T_m 80°C) and is used by several approved antibody-based therapeutics (Na *et al*., 2016). The variable light and heavy chain germ lines present in the Hu4D5-8 Fab framework are V_K1-1 and V_H3-23, respectively, which are frequently seen in human antibody responses (Lee *et al*., 2004B).

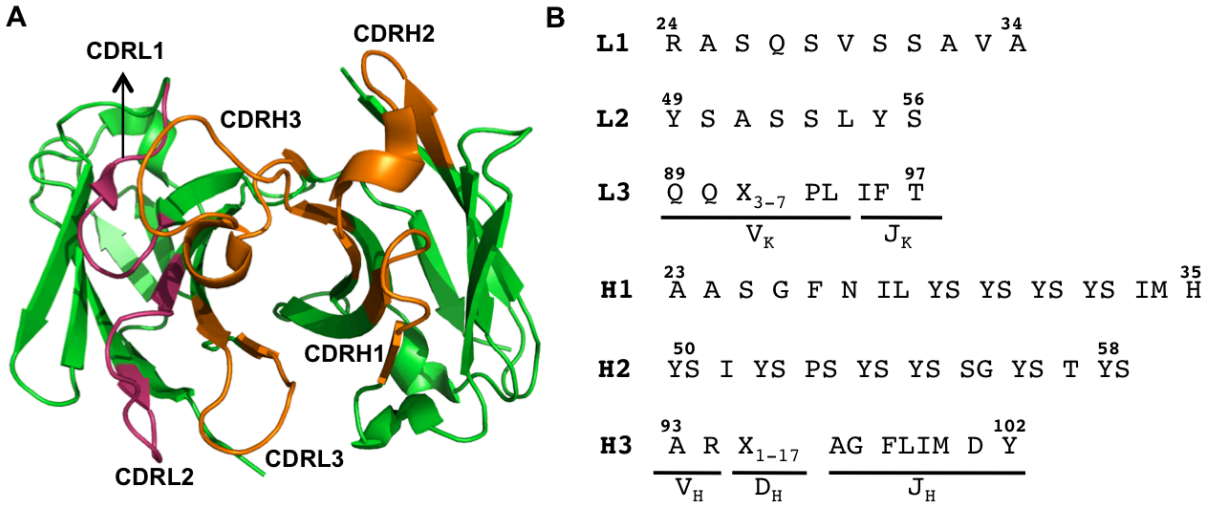


Figure 5.4: Library-F design: (A) Variable domains of the Hu4D5-8 Fab framework used for library construction. The top-down view of anti-MBP Fab (PDB entry: 3PNW) is shown in cartoon representation. Fixed CDRs are colored in pink and diversified CDRs are colored in orange. **(B)** CDR diversity design in library-F. Diversity was restricted to four CDRs (L3, H1, H2 and H3). Amino acids allowed at each CDR position are denoted by the single-letter code. X denotes any of the following nine amino acids introduced at different proportions: Y (25%), S (20%), G (20%), A (10%), F (5%), W (5%), H (5%), P (5%) or V (5%). The lengths of L3 and H3 are varied by altering the number of X. Variable (V), Diversity (D) and Joining (J) segments in CDRs L3 and H3 are indicated. The Kabat scheme is used for numbering amino acids.

Library-F diversity was restricted to three heavy-chain CDRs and CDRL3. In CDRs H1 and H2, solvent-accessible positions were randomized with binary degenerate codons that encode equal proportions of two amino acids, mostly Tyr and Ser. In CDRs L3 and H3, the central region (positions 91-94 encoded by the Vκ gene), and the diversity region (positions 95-99 encoded by the D_H gene), respectively, were randomized using custom-made codon mixes that encoded pre-defined proportions of nine amino acids. The joining segment regions within CDRs L3 and H3 were also softly randomized to resemble the human Jκ4 and J_H4 gene segments, respectively. The library contains 5 different CDRL3 lengths (from 8 to 12 residues) and 17 different CDRH3 lengths (from 7 to 23 residues). The theoretical and actual diversities of library-F are ~10^{28} and ~10^{10}, respectively. The overall design of library-F is outlined in **Figure 5.4**.

Library-S was designed to exist as a simplified and possibly more stable version of Library-F. To quickly prepare samples for NGS analysis, and to easily reconstruct Fabs from NGS information, we included only two diversified CDRs (L3 and H3) in the Hu4D5-8 Fab scaffold whereas library-F contains four diversified CDRs. It is easier and more efficient to link

two diversified CDRs together by Kunkel mutagenesis than linking four CDRs. Also, the CDRL3-CDRH3 pair from library-S selection outputs can be sequenced on 200 bp chips, and reconstruction of rare clones from library-S selections can be accomplished with only two oligonucleotides (for two CDRs). To increase the functional diversity of the library, four fixed CDRs were designed to preserve the most-frequent 'canonical' CDR conformation preferred by the Hu4D5 Fab scaffold. The library diversity was engineered within CDRL3 and CDRH3 using custom-designed trinucleotide phosphoramidite mixes and biased towards human antibody CDR diversities. To compensate for the reduction in CDRH1 and CDRH2 diversities, additional length, amino acid and joining-segment (J_H) diversities were included within CDRH3. Further, to reduce the number of template sequences during library-S construction, we decided to replace the template phagemid CDRs with restriction enzyme sites, and to digest the library DNA after mutagenesis. NGS analysis of the naïve CDRH3 diversity of library-F indicated that the template phagemid used for library construction is retained in ~13% of the population (extended data-Figure 5.17B).

Library-S contains four fixed CDRs (L1, L2, H1 and H2) and two diversified CDRs (L3 and H3). We aligned 100 $V_\kappa 1$ variable light chain sequences and 500 $V_H 3$ variable heavy chain sequences, and used the CDR consensus sequences to fix CDRs L1, L2, H1 and H2 in the Hu4D5-8 Fab framework **(Figure 5.5)**. These CDR sequences correspond to the following canonical CDR conformations described by Chothia and coworkers (Al-Lazikani *et al*., 1997): L1-2A, L2-1, H1-1 and H2-3A, and the following canonical CDR conformations described by North *et al* in 2011: L1-11-1, L2-8-1, H1-13-1 and H2-10-2. We introduced three point mutations (A71R, T73N, and A78L) into the FRM3 region of the Hu4D5-8 Fab framework for proper display of H2-3A/ H2-10-2 CDRH2 canonical loop (Carter *et al*., 1992; Eigenbrot *et al*., 1992; Knappik *et al*., 2000; North *et al*., 2011; Prassler *et al*., 2011; Tiller et al., 2013). The use of pre-defined CDR conformations in library design has been shown to improve the success rate and biophysical properties of synthetic Fab libraries (Rothe *et al*., 2008; Prassler *et al*., 2011; Tiller et al., 2013).

Library diversity was included within CDRs L3 and H3. The CDRL3 length was fixed at nine amino acids. CDRL3 anchor residues (two at the N-terminus and three at the C-terminus of CDRL3) were fixed to favor the L3-9-*cis*7-1 canonical conformation (North *et al*., 2011). The central region (positions 91-94 encoded by the Vκ gene) was diversified using a custom-made

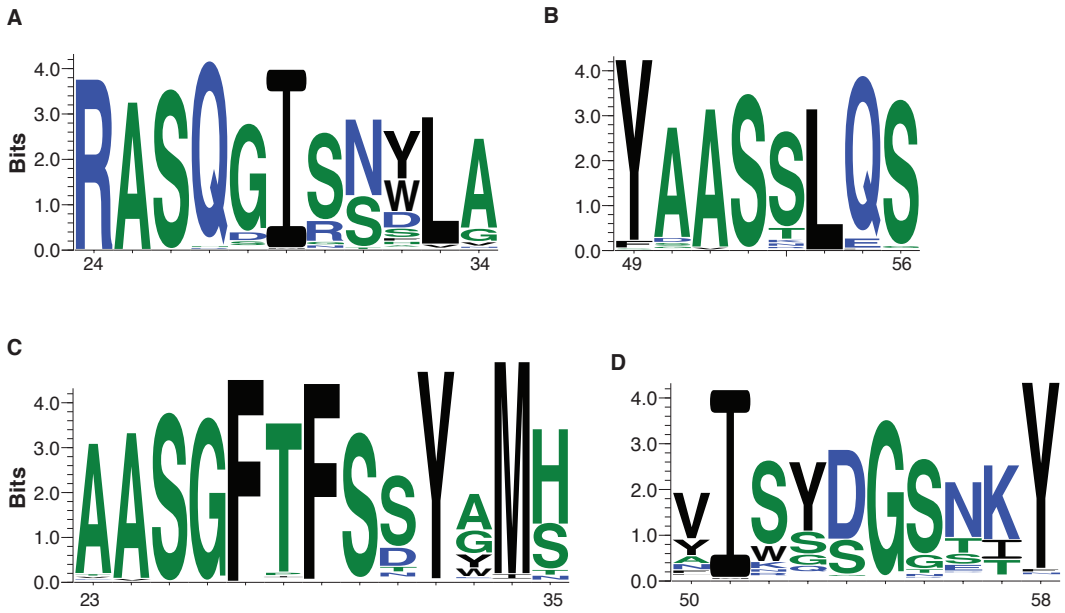


Figure 5.5: Sequence logos used for designing fixed CDRs in library-S. Sequence conservation within **(A)** CDRL1, **(B)** CDRL2, **(C)** CDRH1, and **(D)** CDRH2 regions of Vκ1 (n = 100) and V_H3 (n = 500) antibody sequences. Sequence logos were generated using the WebLogo server. The Kabat scheme was used for numbering amino acids.

codon mix that encoded pre-defined proportions of thirteen amino acids. The theoretical diversity in CDRL3 was 2.9×10^4. Since CDRH3 plays a dominant role in antigen recognition, CDRH3 was designed to contain length, amino acid, conformational, and joining-segment (J_H-region) diversities. 19 different CDRH3 lengths (from 7 to 25 residues) were used in the design. Since an increase in J_H6 gene usage is observed with increasing CDRH3 lengths during the process of human VDJ-recombination (Zemlin *et al*., 2003; Prassler *et al*., 2011), a J_H6 gene segment was incorporated in long CDRH3 sequences (16-25 residues), whereas a J_H4 gene segment was used for short CDRH3 sequences (7-16 residues). In CDRH3, two different codon mixes were used to randomize the diversity region (positions 95-99 encoded by the D_H gene). For short (7-16 residues) and long (16-25 residues) CDRH3 lengths, custom-made codon mixes encoded pre-defined proportions of thirteen and nine amino acids, respectively. To introduce CDRH3 conformational diversity, key anchor residues within J_H4 or J_H6 segments were soft randomized to favor both the bulged and non-bulged CDRH3 conformations (North *et al*., 2011). The theoretical diversity in CDRH3 was 2.6×10^{14}. The theoretical diversity of Library-S was 7.4×10^{18} members. The overall design of library-S is outlined in **Figure 5.6**.

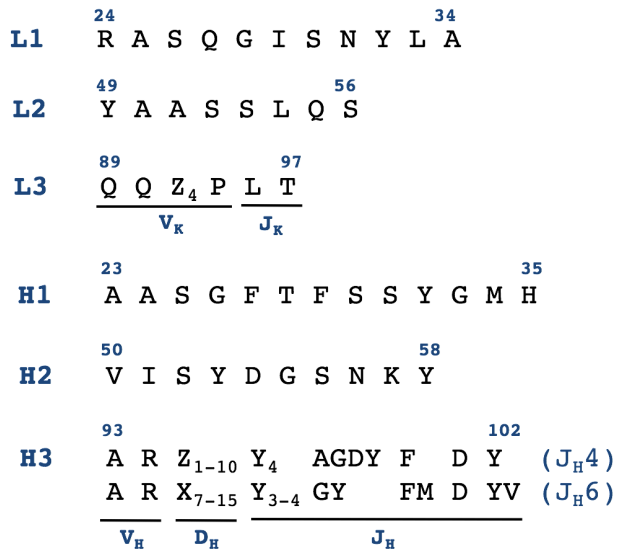


Figure 5.6: Library-S design. Library-S contains four fixed CDRs (L1, L2, H1, and H2) and two diversified CDRs (L3 and H3). Z denotes any of the following thirteen amino acids introduced at different proportions: Y (20%), S (20%), G (20%), T (6.5%), A (6.5%), P (6.5%), H (3.5%), R (3.5%), E (3.5%), F (2.5%), W (2.5%), V (2.5%) or L (2.5%). X denotes any of the following nine amino acids introduced at different proportions: Y (25%), S (20%), G (20%), A (10%), F (5%), W (5%), H (5%), P (5%) or V (5%). CDRH3 length is varied by altering the number of X and Z. Variable (V), Diversity (D) and Joining (J) segments in CDRs L3 and H3 are indicated. The Kabat scheme is used for numbering amino acids.

Hu4D5-8 is based on the anti-HER2 mouse monoclonal antibody (Mu4D5). Humanization of Mu4D5 V_L and V_H domains required 24 and 36 amino acid changes, respectively. The fully humanized antibody (Hu4D5-1) binds to HER2 with 80-fold lower affinity than Mu4D5 and does not inhibit HER2-overexpressing cells. Two positions in V_L (E55Y and G66R) and five positions in V_H domain (R71A, D73T, L78A, A93S and V102Y) were reverted to amino acids present in Mu4D5 to improve binding. These substitutions located in CDR anchor residues (positions 55^{VL}, 93^{VH} and 102^{VH}) and framework residues (positions 66^{VL}, 71^{VH}, 73^{VH} and 78^{VH}), which are known to critically affect the conformation of CDRs. HU4D5-8 binds 3-fold stronger than Mu4D5, and has anti-proliferative activity comparable to Mu4D5 (Carter *et al*., 1992; Kelley *et al*., 1992; Eigenbrot *et al*., 1992). Relative to Hu4D5-8, which retains seven mouse residues, the 4D5 framework used in library-S contains only one mouse residue (R66 in V_L). The 4D5 framework used in library-F contains four mouse residues (R66 in V_L, and A71, T73 and A78 in V_H). To highlight the amino acid differences between various 4D5 frameworks, a multiple sequence alignment is included in **Figure 5.7**.

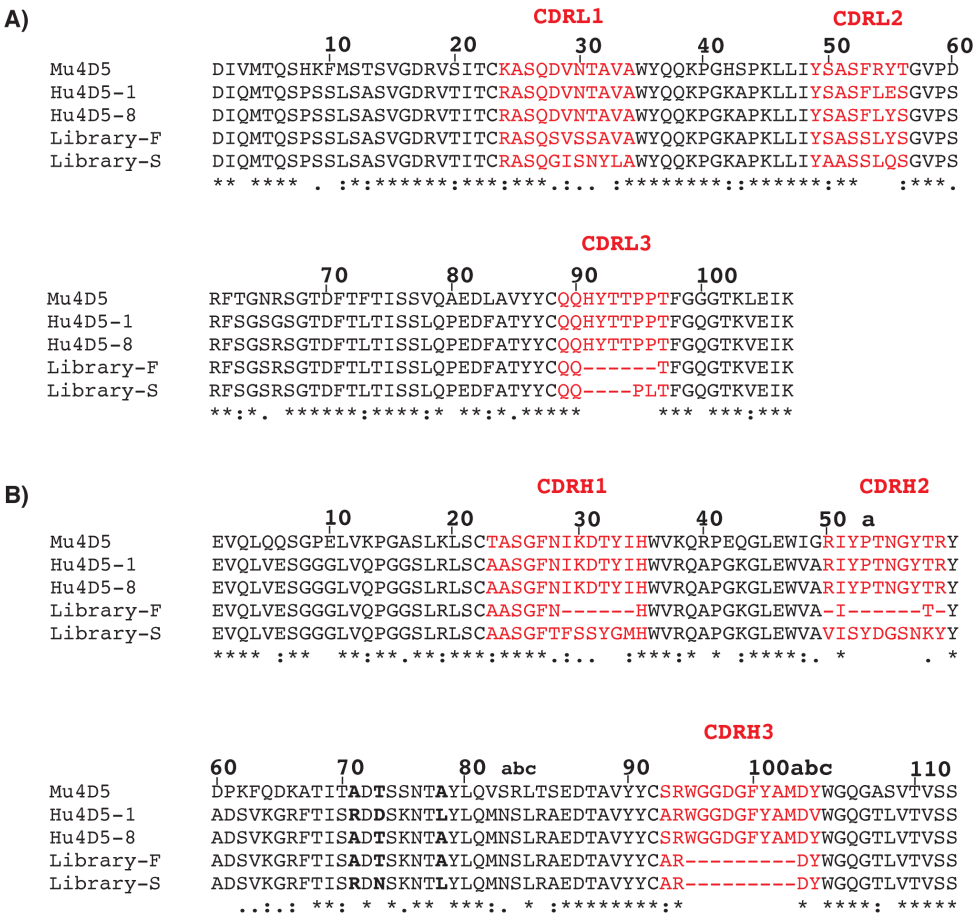


Figure 5.7: Multiple sequence alignment of 4D5 framework sequences: **(A)** the variable light chain region and **(B)** the variable heavy chain region. The following sequences were included in the alignment: anti-HER2 murine 4D5 Fab (Mu4D5), Hu4D5-1 (fully humanized form of Mu4D5), Hu4D5-8 (humanized form of Mu4D5 and Herceptin Fab), Library-F (4D5 framework used in library-F) and Library-S (4D5 framework used in library-S). Hyphen indicates diversified positions in library-F and library-S. Symbols asterisk, colon, period and space indicate identical residues, conservative substitutions, semi-conservative substitutions and mismatches, respectively. Three V_H substitutions (71, 73 and 78) in the FRM3 region are shown in bold letters. CDR residues are colored in red. Sequences were aligned using the Clustal Omega program. The Kabat scheme is used for numbering amino acids. CDRs were defined according to North *et al*., 2011.

5.1.3 Library-S: Construction and Quality Control

Library-S was constructed using an established M13 bacteriophage system that allows bivalent Fab display (Lee *et al*., 2004A). A Hu4D5-8 Fab-encoding phagemid whose CDRs have been replaced with NotI sites was used as a template to construct libraries. Kunkel mutagenesis was used to repair the NotI sites and replace CDR positions with fixed or degenerate codons

encoding the amino acid composition shown in Figure 5.6. Upon digestion of the non-mutated template DNA by NotI, library DNA from the mutagenesis reaction was electroporated into an *E. coli* strain suitable for high-efficiency transformation and phage production. To minimize biases in bacterial growth and phage production due to differences in CDRH3 lengths, we constructed library-S in 20 different reactions, each reaction representing a CDRH3 length, and rescued 20 sub-libraries (SL1 - SL20) separately.

To assess the quality of sub-libraries, we performed a NGS analysis of their naïve CDRH3 diversity. On average, we obtained ~1,000 filtered-reads per pool, and determined their overall sequence composition **(Figure 5.8A)**. Except for SL13, SL15 and SL20, the proportion of incorrect sequences (sum of out-of-frame sequences, template sequences and sequences with stop codons) in the sub-libraries was less than 10%. The average proportion of incorrect sequences present in 20 sub-libraries was only 5.7%. Satisfied with the quality of sub-libraries, we combined equal number of phages from 20 sub-libraries to make library-S. To assess library-S quality, we sequenced both the diversified CDRs in the library. We obtained 13,118 CDRL3 and 13,712 CDRH3 sequences, and determined their overall sequence composition. The proportion of incorrect CDRL3 and CDRH3 sequences was only 1.8% and 3.4%, respectively. Since the library size was higher than the theoretical CDRL3 diversity, the proportion of unique CDRL3 sequences was only 40%. Also, for a few sub-libraries with short CDRH3 lengths, the library size was higher than the theoretical CDRH3 diversity, resulting in a decrease in the proportion of unique CDRH3 sequences to 81%. Considering the randomness in the pairing of CDRs L3 and H3, the actual proportion of unique Fabs would increase in the library.

To assess library-S diversity in detail, we linked the two diversified CDRs together, sequenced the CDRL3-CDRH3 strip on a 200 bp chip, and obtained 188,529 sequences. Template retention in the library was only 2%, indicating that restriction enzyme digestion of the library mutagenesis reaction is an efficient strategy to reduce the amount of template sequences. Sequence analysis showed the presence of 94% unique sequences, and 91% sequences with a single copy in the library. Further, the proportion of the most frequent CDRH3 sequence in the library was only 0.02%, which indicated that there is no significant sequence-bias in the library. After electroporation of library DNA into *E. coli*, we estimated the size of the library to be 9.4 billion Fabs by bacterial titrations. Having determined the proportion of correct and unique sequences by deep sequencing, we estimated that there are 8.5 billion unique Fabs in the library.

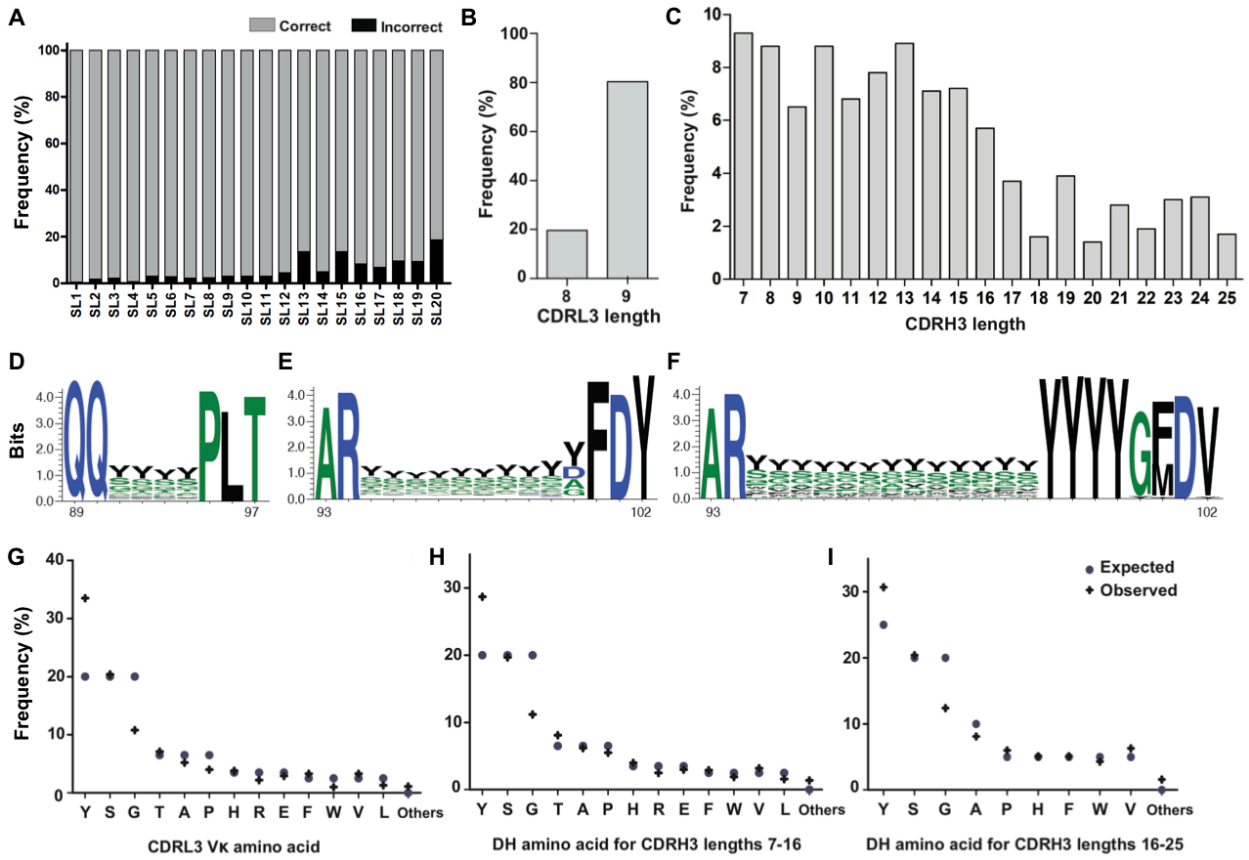


Figure 5.8: NGS analysis of the naïve library-S diversity. (A) Overall composition of CDRH3 sequences in sub-libraries (SL) 1-20. **(B)** CDRL3 length distribution. **(C)** CDRH3 length distribution. Sequence logos showing the positional amino acid composition of **(D)** CDRL3 sequences containing 9 residues, **(E)** J_H4 CDRH3 sequences containing 15 residues and **(F)** J_H6 CDRH3 sequences containing 23 residues. **(G)** Total amino acid composition of the CDRL3 central region. **(H)** Total amino acid composition of the CDRH3 D_H segment for lengths 7-16. **(I)** Total amino acid composition of the CDRH3 D_H segment for lengths 16-25. The expected amino acid composition (encoded by phosphoramidite trimers) is also included (filled circles).

Next, we analyzed CDRL3 and CDRH3 sequences in terms of length distribution. Although CDRL3 length was fixed to 9 residues, we observed two different CDRL3 lengths in the population **(Figure 5.8B)**. 20% of CDRL3 sequences had only 8 residues. One of the four diversified positions in the CDRL3 central region was found to be absent. A possible explanation for this deletion could be a DNA synthesis error that could have happened during the incorporation of phosphoramidite codons into oligonucleotides. Considering that the Hu4D5-8 framework can accommodate CDRL3 lengths ranging from 8 to 12 residues (Persson *et al*., 2013), we did not try to eliminate the CDRL3 sequences with 8 residues. With respect to

CDRH3, 19 loop lengths were included in the design: lengths 7-16 had a J_H4 segment and lengths 16-25 had a J_H6 segment. Since we mixed equal number of phages from 20 sub-libraries, we expected to obtain a flat CDRH3 length distribution. However, the analysis showed that J_H4-containing CDRH3 sequences (74%) were overrepresented relative to J_H6-containing CDRH3 sequences (26%) **(Figure 5.8C)**.

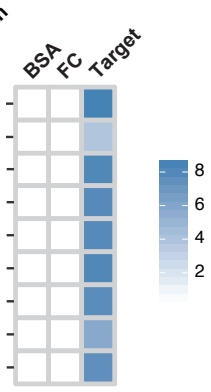
Next, we analyzed the amino acid composition of CDRL3 and CDRH3. Sequence logo analysis showed a clear difference in sequence conservation between CDR anchor residues and central variable positions. Also, sequence logos indicated that CDR diversification was uniform across variable positions and no major deviations were seen between the expected and observed amino acid compositions of anchor residues. Representative sequence logos are shown in **Figures 5.8D, 5.8E and 5.8F.** To obtain a more refined amino acid distribution, we analyzed the composition of 13 possible amino acids within the CDRL3 central region **(Figure 5.8G)**, 13 possible amino acids within the D_H segment of J_H4-containing CDRH3 sequences (**Figure 5.8H)**, and 9 possible amino acids within the D_H segment of JH6-containing CDRH3 sequences (**Figure 5.8I).** These regions have the highest diversity in the library and are encoded by one of two codons X and Z with predefined proportions of 9 and 13 amino acids, respectively. While the frequency of most residues was close to frequencies predefined in codons X and Z, tyrosine was overrepresented and glycine was underrepresented compared to the design.

5.1.4 Library-S Validation

To verify library-S functionality, we conducted four rounds of solid-phase selections against two human cell-surface receptor ectodomains Jagged-1 and Jagged-2. Briefly, we immobilized recombinant target-Fc fusion proteins on Maxisorp plates, and incubated target-coated wells with library-S deselected for binding to the Fc protein. Upon elimination of non-specific phages by washing, we eluted bound phages and amplified them in *E. coli* overnight for subsequent rounds of panning. Random clone picking and Sanger sequencing of ten phagemids from round-3 and round-4 phage pools identified two sequence-unique clones for Jagged-1 and seven sequence-unique clones for Jagged-2. Phage-ELISA showed that all Jagged-1 and Jagged-2 Fab clones bound to the target receptor but not to BSA and the Fc protein **(Figure 5.9A)**. We then subcloned two Jagged-1 Fab clones and four Jagged-2 Fab clones into a Fab expression vector and expressed and purified Fabs from *E. coli* using Protein-L affinity chromatography. We observed high Fab expression rates (estimated average ~17.4 mg/L), suggesting that new

A

Target	Clone	CDRL3	CDRH3	H3 length
Jagged-1	J1/S/1	QQYLGLPLT	ARWSPWGVYGFDY	13
Jagged-1	J1/S/2	QQSLATPLT	ARSEYGTAGYFDY	13
Jagged-2	J2/S/1	QQGEYTPLT	ARGSYYVPAFDY	12
Jagged-2	J2/S/2	QQGSYSPLT	ARGHYTTPDFDY	12
Jagged-2	J2/S/3	QQGSFGPLT	ARGHYLFPDFDY	12
Jagged-2	J2/S/4	QQGTYLPLT	ARGYYSYPGFDY	12
Jagged-2	J2/S/5	QQGGHSPLT	ARPSYSYGSFDY	12
Jagged-2	J2/S/6	QQSSYSPLT	ARGHYFRRGFDY	12
Jagged-2	J2/S/7	QQGSYSPLT	ARGFFYYPAFDY	12



B

Target	Clone	Fab yield (mg/L)	EC_{50} (nM)	K_D (nM)
Jagged-1	J1/S/1	18.7	5.88 ± 2.34	0.28 ± 0.01
Jagged-1	J1/S/2	21.5	40.3 ± 2.69	0.48 ± 0.04
Jagged-2	J2/S/1	14.6	1.23 ± 0.17	0.17 ± 0.01
Jagged-2	J2/S/2	14.6	1.32 ± 0.19	0.69 ± 0.01
Jagged-2	J2/S/4	18.8	1.41 ± 0.12	0.08 ± 0.01
Jagged-2	J2/S/5	16.0	54.96 ± 0.23	3.81 ± 0.07

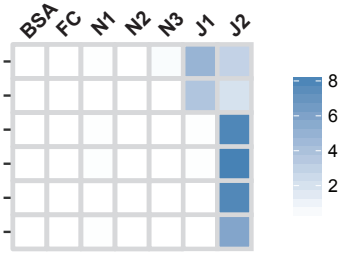


Figure 5.9: Jagged-1 and Jagged-2 Fabs from library-S. (A) Sequence characteristics and phage-displayed Fab binding characteristics for two Jagged-1 Fabs and seven Jagged-2 Fabs isolated by Sanger sequencing. **(B)** Fab yield, affinity and specificity of purified Jagged-1 and Jagged-2 Fabs. Fab yield was estimated for 1 L bacterial culture. EC_{50} values and K_D values for Fabs binding to their intended targets were determined by multi-point Fab-ELISA and bio-layer interferometry, respectively. Fab specificity was determined using single-point Fab-ELISA at 100 nM Fab concentration. Phage-ELISA and specificity-ELISA ABS_{450} values are shown as heatmaps. In ELISA heatmaps, Notch-1/2/3 and Jagged-1/2 target proteins are indicated as N1, N2, N3, J1 and J2.

CDRs and framework mutations are well accommodated by the Fab-4D5 framework. Next, we determined EC_{50} values for Fabs binding to Jagged-1 or Jagged-2 using Fab-ELISA. J1/S/1 bound to Jagged-1 with an EC_{50} of 6 nM. J1/S/2 bound ~7-fold weaker than J1/S/1 (EC_{50} = 40 nM). J2/S/1, J2/S/2 and J2/S/4 bound to Jagged-2 with an EC_{50} value of ~1 nM, and J2/S/5 had an EC_{50} value of 50 nM (**Figure 5.9B)**. Next, we measured the kinetics of Fab binding using bio-layer interferometry. Briefly, Fabs were immobilized on amine-reactive sensors and sensor tips were exposed to increasing concentrations of target receptors. Association and dissociation rates were assessed by a wavelength shift and kinetic data sets were fit using a 1:1 binding model. Both Jagged-1 Fabs had sub-nM K_D values: 0.3 nM for J1/S/1 and 0.5 nM for J1/S/2. Among the four Jagged-2 Fabs, J2/S/4 was the highest-affinity binder (K_D = 80 pM), and J2/S/5 was the

lowest-affinity binder (K_D = 3.8 nM). J2/S/1 and J2/S/2 possessed sub-nM K_D values, 0.17 nM and 0.69 nM, respectively (**Figure 5.9B)**. In total, all the six Fabs bound very strongly to their targets with EC_{50} values ranging from 1 nM to 50 nM, and K_D values ranging from 80 pM to 3.8 nM. Next, to assess Fab specificity, we tested the binding of Fabs to Notch and Jagged receptor ectodomains using Fab-ELISA. At 100 nM Fab concentration, Jagged-1 Fabs cross-reacted with Jagged-2, whereas Jagged-2 Fabs exhibited target-specific binding **(Figure 5.9B)**. Kinetic analysis showed that Jagged-1 Fabs, J1/S/1 and J1/S/2, bound to Jagged-2 with K_D values of 3.2 ± 0.1 nM and 7.2 ± 0.3 nM, respectively. Jagged-1 Fabs showed ~10-fold specificity over Jagged-2, and Jagged-2 Fabs showed >1000-fold specificity over Jagged-1. Together, these results highlighted the potential of library-S in generating Fabs with very high affinity and specificity against cell-surface receptors.

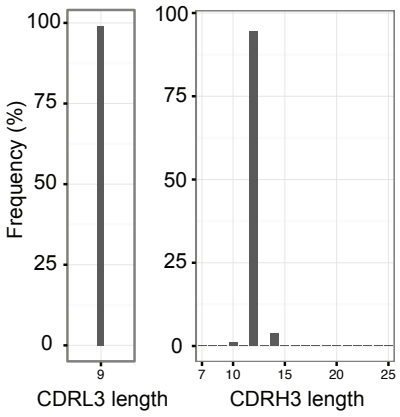
5.1.5 Validation of the NGS-assisted Antibody Reconstruction Platform

To demonstrate the feasibility of our Fab reconstruction approach, we used the Jagged-2 selection output. Like others before (Ravn *et al*., 2010; Venet *et al*., 2012), we observed a significant enrichment in target-specific CDRH3 lengths and sequences after the third round of selection (extended data, Figures 5.13C, 5.15C and 5.15F); therefore, we chose to use the round-3 phage pool for CDRL3-CDRH3 strip sequencing. We generated and sequenced the CDR strip as illustrated in Figure 5.2, and ranked the CDRL3-CDRH3 sequences based on their relative frequencies. A list of enriched CDR-combinations (with frequencies >1%) is shown in **Figure 5.10A**. The seven CDRL3-CDRH3 sequences identified by random clone picking and Sanger sequencing identically matched to seven CDR-combinations in the list. Next, we parsed the CDRL3 and CDRH3 regions, and analyzed them in terms of length distribution **(Figure 5.10B)**. In agreement with Sanger sequencing results, a significant enrichment was observed for CDRL3 length of 9 residues and CDRH3 length of 12 residues. Also, a minor enrichment was noted with two CDRH3 lengths (10 and 14 residues). We then extracted the sequences containing CDRH3 length of 12 residues, and generated sequence logos for CDRL3 and CDRH3 regions **(Figure 5.10C)**. Enrichment of specific amino acids was observed in the diversified regions of both CDRs. Notably, Gly91 in CDRL3, and Gly95, Tyr97 and Pro100 in CDRH3. J2/S/5 had mismatches in two of these positions (Pro95 and Gly100) and bound weaker than other three Fab clones. Next, to assess the binding properties of clones corresponding to less frequent CDRH3 lengths, we reconstructed three rare clones with CDRH3 lengths of 10 or 14 residues. J2/S/R1

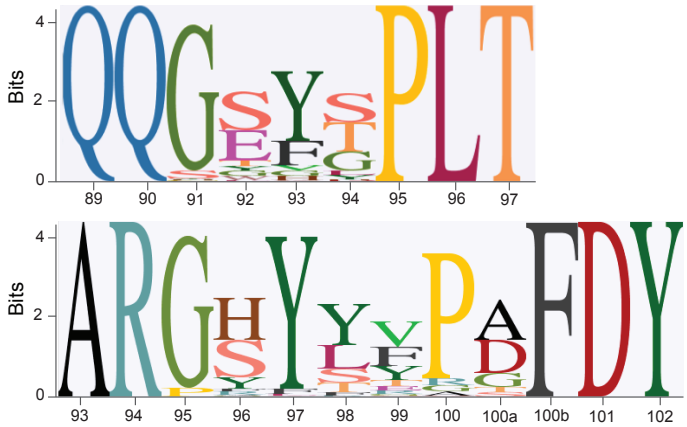
A

Rank	CDRL3-CDRH3	Round 3 (%)	
1	QQGEYTPLT-ARGSYYVPAFDY	25.41	← J2/S/1
2	QQGSFGPLT-ARGHYLFPDFDY	18.18	← J2/S/3
3	QQGSYSPLT-ARGHYTTPDFDY	4.65	← J2/S/2
4	QQGYVSPLT-ARGHYYEPTFDY	3.84	
5	QQGTYLPLT-ARGYYSYPGFDY	3.43	← J2/S/4
6	QQGGHSPLT-ARPSYSYGSFDY	2.95	← J2/S/5
7	QQGSYSPLT-ARGYYSYPGFDY	2.80	
8	QQGWGYPLT-ARVASPYWGYDFDY	2.76	
9	QQSSYSPLT-ARGHYFRRGFDY	2.45	← J2/S/6
10	QQGSYSPLT-ARGFFYYPAFDY	2.05	← J2/S/7
11	QQGSYSPLT-ARGRYTGAAFDY	1.92	

B

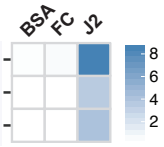


C



D

Clone	CDRL3	CDRH3	H3 length	Round 3 (%)
J2/S/R1	QQGEYHPLT	ARPYVADFDY	10	0.42
J2/S/R2	QQSSYTPLT	ARAPFRYFDY	10	0.03
J2/S/R3	QQGWGYPLT	ARVASPYWGYDFDY	14	2.76



E

Clone	Fab yield (mg/L)	EC_{50} (nM)	K_D (nM)
J2/S/R2	9.4	199.8 ± 3.98	5.06 ± 0.13
J2/S/R3	12.8	2.23 ± 0.11	NA

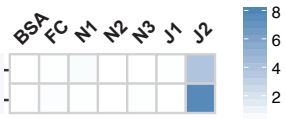
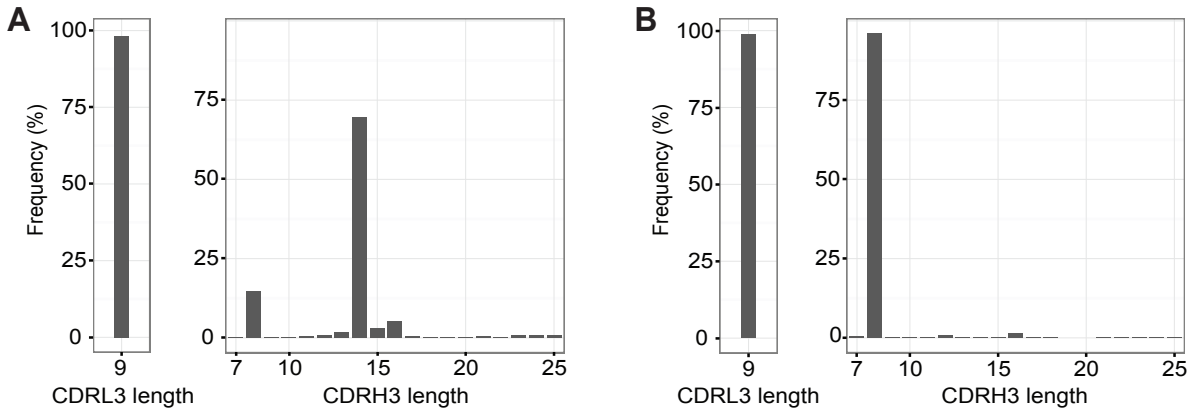


Figure 5.10: NGS-assisted reconstruction of rare Fab clones from the Jagged-2 selection. **(A)** Enriched sequences (> 1%) from the round-3 CDRL3-CDRH3 strip sequencing output. Seven Fabs isolated by Sanger sequencing are indicated. **(B)** CDRL3 and CDRH3 length distribution of the Jagged-2 selection output (round-3). **(C)** Sequence logos showing the positional amino acid composition of CDRL3 and CDRH3 sequences from the Jagged-2 selection output. **(D)** Sequence characteristics and phage-displayed Fab binding characteristics for three Jagged-2 rare clones reconstructed from CDRL3-CDRH3 NGS information. **(E)** Fab yield, affinity and specificity of purified rare Fabs J2/S/R2 and J2/S/R3. EC_{50} and K_D values were determined by multi-point Fab-ELISA and bio-layer interferometry, respectively. Fab specificity was determined using single-point Fab-ELISA at 100 nM Fab concentration. Phage-ELISA and specificity-ELISA ABS_{450} values are shown as heatmaps.

and J2/S/R2 had short CDRH3 lengths (10 residues). J2/S/R1 had a frequency of 0.4%, and showed strong binding in phage-ELISA. J2/S/R2 had a frequency of 0.03%, and showed moderate binding in phage-ELISA **(Figure 5.10D)**. The purified Fab J2/S/R2 bound to Jagged-2 with EC_{50} and K_D values of 200 nM and 5.06 nM, respectively. J2/S/R2 also had strict specificity for Jagged-2. J2/S/R3 had a CDRH3 length of 14 residues, and had the highest frequency (2.8%) among rare clones. The EC_{50} of the purified Fab J2/S/R3 (2.23 nM) was similar to the EC_{50} values of top clones. As expected, J2/S/R3 did not cross-react with other targets in Fab-ELISA **(Figure 5.10E)**. Together, these results indicated that less frequent clones with low-nM affinity and high specificity could be reconstructed from phage selection outputs.

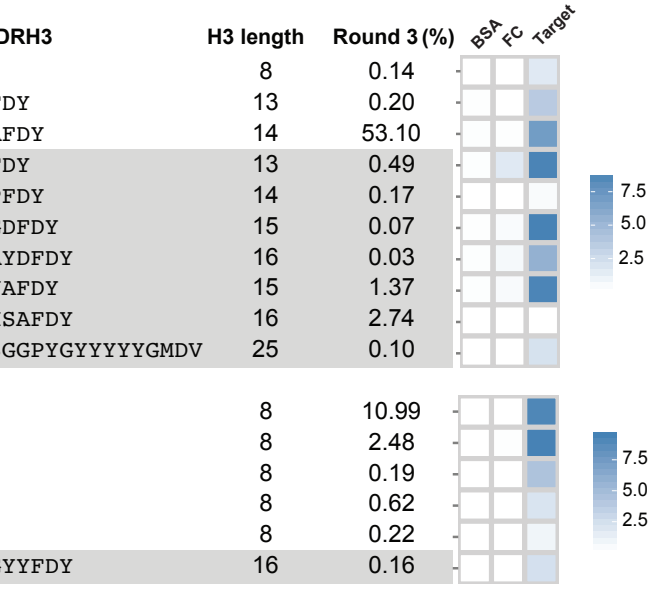
5.1.6 Reconstruction of Rare Fab clones with Better Binding Properties

In case of selections against Notch-2 and Notch-3, top clones isolated by Sanger sequencing had low affinity and/or specificity, therefore we used our Fab reconstruction approach to check whether rare clones with better binding properties can be recovered from selection outputs. For both targets, we subjected the round-3 phage pool to CDRL3-CDRH3 strip sequencing and analyzed the sequences in terms of relative abundance and length distribution. Sequence analysis, phage and Fab work for both the targets are summarized in **Figure 5.11**. For the Notch-2 selection, Sanger sequencing gave rise to three unique Fab clones. While N2/S/3 had the highest frequency (53.1%) in the round-3 phage pool, N2/S/1 and N2/S/2 had very low frequencies (~0.2%). N2/S/1 and N2/S/3 performed poorly in Fab-ELISA and kinetic analyses. Despite its low abundance, N2/S/2 bound to Notch-2 with a low-nM K_D (5 nM). However, N2/S/3 cross-reacted with Notch-3 and Jagged-2 **(Figure 5.11D)**. Sequence analysis indicated the enrichment of three types of Fab sequences for the Notch-2 selection: (I) sequences with CDRH3 lengths of eight residues (similar to N2/S/1); (II) sequences containing a leucine residue in the first diversified position of CDRH3 (similar to N2/S/2) and (III) sequences containing the FGG-motif in the N-terminus of CDRH3 (similar to N2/S/3). With CDRH3 length of eight residues, type-I sequences contain five fixed anchor residues and only three diversified positions. Since we have observed the enrichment of type-I sequences in many unsuccessful selections (data not shown), we did not reconstruct them. The combination of short CDRH3 length and very low CDRH3 diversity might have resulted in their non-specific behavior. Type-II and type-III sequences were found in four different CDRH3 lengths (13, 14, 15 and 16 residues for type-II, and 14, 15, 16 and 25 residues for type-III). We reconstructed the most-frequent Fab sequence



C

Clone	CDRL3	CDRH3	H3 length	Round 3 (%)
N2/S/1	QQRYSYPLT	ARSFGFDY	8	0.14
N2/S/2	QQSYTSPLT	ARLASYRAPDFDY	13	0.20
N2/S/3	QQRYAYPLT	ARFGGWTAYAAFDY	14	53.10
N2/S/R1	QQSTRVPLT	ARLSGYYEGDFDY	13	0.49
N2/S/R2	QQGSEVPLT	ARLYTSYYGYPFDY	14	0.17
N2/S/R3	QQSYTWPLT	ARLALPPYHTGDFDY	15	0.07
N2/S/R4	QQGYYYPLT	ARLWGYTPGSAYDFDY	16	0.03
N2/S/R5	QQRSVYPLT	ARFGGHGSFTYAFDY	15	1.37
N2/S/R6	QQRYHSPLT	ARFGGYSRVPHSAFDY	16	2.74
N2/S/R7	QQRVYYPLT	ARFGGGGSSYSGGPYGYYYYYGMDV	25	0.10
N3/S/1	QQEHSWPLT	ARSFAFDY	8	10.99
N3/S/2	QQGYTVPLT	ARSYGFDY	8	2.48
N3/S/3	QQAYYYPLT	ARSFGFDY	8	0.19
N3/S/4	QQGYSFPLT	ARRFAFDY	8	0.62
N3/S/5	QQARSFPLT	ARHFGFDY	8	0.22
N3/S/R1	QQYYWAPLT	ARSRYSPYAYGYYFDY	16	0.16



D

Clone	Fab yield (mg/L)	EC_{50} (nM)	K_D (nM)
N2/S/1	13.9	> 1000	109.4 ± 7.33
N2/S/2	18.3	78.93 ± 22.02	5.05 ± 0.31
N2/S/3	4.9	> 1000	64.58 ± 2.81
N2/S/R7	14.6	15.93 ± 2.29	7.68 ± 0.72
N3/S/1	13.1	> 1000	37.51 ± 4.95
N3/S/2	15.3	> 1000	60.6 ± 2.51
N3/S/R1	11.5	16.04 ± 1.12	2.65 ± 0.27

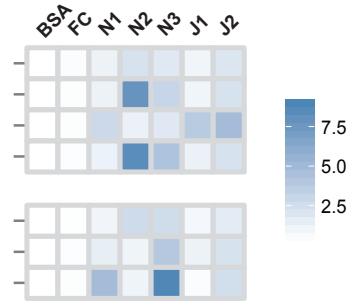
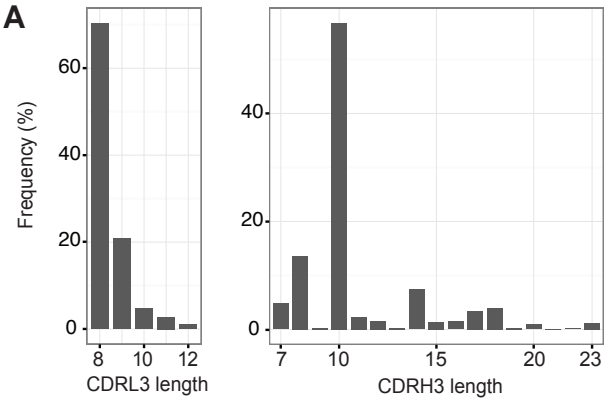


Figure 5.11: Notch-2 and Notch-3 Fabs from Library-S. CDRL3 and CDRH3 length distributions of **(A)** Notch-2 and **(B)** Notch-3 selection outputs. **(C)** Sequence characteristics and phage-displayed Fab binding characteristics for Notch-2 Fabs (top panel) and Notch-3 Fabs (bottom panel). **(D)** Fab yield, affinity and specificity of Notch-2 Fabs (top panel) and Notch-3 Fabs (bottom panel). In panels C and D, reconstructed rare clones are highlighted in gray. Fab specificity was determined at 1 μM Fab concentration. Phage-ELISA and specificity-ELISA ABS_{450} values are shown as heatmaps.

from each enriched CDRH3 length for both types II and III, and tested them for binding to Notch-2 by phage-ELISA. Among the type-II Fabs, N2/S/R3 and N2/S/R4 showed higher ELISA output than N2/S/2, N2/S/R1 cross-reacted with the Fc protein, and N2/S/R2 did not bind to Notch-2. Among the type-III Fabs, N2/S/R5 showed higher ELISA output than N2/S/3, N2/S/R7 showed lower ELISA output than N2/S/3, and N2/S/R6 did not bind to Notch-2 **(Figure 5.11C)**. In total, four out of seven rare clones tested positive in phage-ELISA. We purified one of the four ELISA-positive clones (N2/S/R7) that had the longest CDRH3 length and assessed its affinity and specificity for Notch-2. N2/S/R7 had EC_{50} and K_D values of 15.94 nM and 7.68 nM, respectively. Similar to N2/S/2, N2/S/R7 cross-reacted with Notch-3 and Jagged-2 **(Figure 5.11D)**. Even though N2/S/R7 had the lowest frequency among the type-III sequences (0.1%), it bound ~8-fold stronger than N2/S/3 that had the highest frequency (53.1%). For the Notch-3 selection, Sanger sequencing gave rise to five clones that matched with the description for type-I sequences. In phage-ELISA, two clones (N3/S/1 and N3/S/2) showed strong binding to Notch-3 whereas three clones (N3/S/3, N3/S/4 and N3/S/5) showed moderate to weak binding **(Figure 5.11C)**. When phage-displayed Fabs were converted into soluble Fabs, N3/S/1 and N3/S/2 showed weak binding to Notch-3 in Fab-ELISA and kinetic analyses **(Figure 5.11D)**. CDRH3 length analysis confirmed the enrichment of type-I sequences in the Notch-3 selection and in addition, showed a minor enrichment for sequences with 16 residues. We reconstructed the most-frequent Fab sequence (0.16%) from CDRH3 length of 16 residues, and assessed its affinity and specificity for Notch-3. N3/S/R1 had EC_{50} and K_D values of 16.04 nM and 2.65 nM, respectively. In Fab-ELISA, N3/S/R1 cross-reacted with Notch-1 and Jagged-2 **(Figure 5.11D)**. For both Notch-2 and Notch-3 selections, rare clones reconstructed from NGS information bound better than top clones isolated by Sanger sequencing. Rare Fabs N2/S/R7 and N3/S/R1 both possessed low-nM K_D values for their targets. However, both Fabs cross-reacted with two other Notch/Jagged receptors.

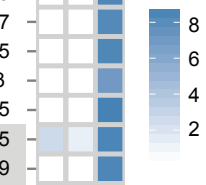
5.1.7 Reconstruction of Rare Fab clones from Library-F Selections

To demonstrate the feasibility of reconstructing Fabs with four diversified CDRs, we chose Jagged-2 as a target. We panned library-F against the Jagged-2 ectodomain, and subjected the round-3 phage pool to L3-H1-H2-H3 strip sequencing as illustrated in Figure 5.2. We parsed the CDRL3 and CDRH3 regions, and analyzed them in terms of length distribution **(Figure 5.12A)**. A significant enrichment was observed for CDRL3 sequences with 8 residues (70.4%)



B

Clone	CDRL3	CDRH1	CDRH2	CDRH3	L3 length	H3 length	Round 3 (%)
J2/F/1	QQFAGYLIT	LSSSYMH	SISSSYGSTS	ARHALDY	9	7	1.86
J2/F/2	QQFYYGLIT	LSYSYMH	SISSSSGYTS	ARYALDY	9	7	0.55
J2/F/3	QQYWGALIT	IYYSYIH	SIYSSYGYTY	ARGPAMDY	9	8	7.36
J2/F/4	QQVVYYGGLIT	LSYYSMH	SIYSYYGYTY	ARSYGLDY	11	8	1.06
J2/F/5	QQWWHPIT	LYSSSMH	SIYSSYGSTY	ARSGFWAMDY	8	10	45.9
J2/F/6	QQSWYYHLIT	LSYYYMH	SISPYYGSTY	ARWSSWAMDY	10	10	1.17
J2/F/7	QQYFYPIT	LYYSSMH	YISPYSSYTY	ARYGYAPGGHGMDY	8	14	5.65
J2/F/8	QQAYVPFT	IYSSSMH	YISPSYSYTY	ARGYSYHYYWAYWGLDY	8	17	2.3
J2/F/9	QQVYYAPIT	IYYSSMH	YISSSYGYTS	ARYGWHWGHSSHSAGLDY	9	18	1.25
J2/F/R1	QQWYGYLIT	FSSSSIH	SIYPYYGYTY	ARSYWPWAFDY	9	11	0.65
J2/F/R2	QQYWAGWWLFT	FSSSSIH	SISSSYGYTY	ARGYYWAPAMDY	11	12	0.29
J2/F/R3	QQWVAPIT	LYYYSMH	SIYPYYGYTY	ARYAPAAVYFVGMDY	8	15	0.48
J2/F/R4	QQWHHPPAPIT	LYSSYIH	YISPSYGYTY	ARYYYSGWAVYAGMDY	11	16	0.3
J2/F/R5	QQSYWPIT	LYSYSMH	YISPSYSYTY	ARYWSAHWSWAHSGGMDY	8	18	0.45
J2/F/R6	QQSYASVLIT	LSYSSMH	YISPSYGYTS	ARAHWSSYVSGGWAYPHHSGMDY	10	23	0.55
J2/F/R7	QQYASYPIT	IYSSSMH	YISPSYSYTY	ARSYWYSHYAWWWASYSVVGIDY	9	23	0.1



C

Clone	Fab yield (mg/L)	EC_{50} (nM)
J2/F/1	3.1	6.03 ± 0.7
J2/F/3	9.8	6.6 ± 1.2
J2/F/5	18.2	155 ± 32
J2/F/6	4.3	1.93 ± 0.76
J2/F/7	8.5	3.11 ± 0.3
J2/F/8	15.2	~ 1000
J2/F/9	4.1	2.99 ± 0.91
J2/F/R2	5	2.38 ± 1.2
J2/F/R3	5.6	1.66 ± 0.71
J2/F/R4	5.6	2.65 ± 1.11
J2/F/R5	6	3.45 ± 0.4
J2/F/R6	8.7	3.72 ± 0.83
J2/F/R7	10	24.62 ± 1.56

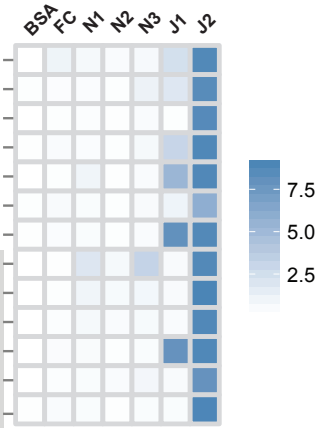


Figure 5.12: Jagged-2 Fabs from library-F. (A) CDRL3 and CDRH3 length distribution of the Jagged-2 selection output (round-3). **(B)** Sequence characteristics and phage-displayed Fab binding characteristics of 9 top Fabs isolated by Sanger sequencing and 7 rare Fabs reconstructed from L3-H1-H2-H3 NGS information. **(C)** Fab yield, affinity and specificity of purified Fabs. Fab specificity was determined at 1 μM Fab concentration. Phage-ELISA and specificity-ELISA ABS_{450} values are shown as heatmaps. In panels B and C, reconstructed rare clones are highlighted in gray.

and CDRH3 sequences with 10 residues (56.8%). CDRH3 sequences with 7, 8 and 14 residues were present between 5-10%, and CDRH3 sequences with 11, 12, 15, 16, 17, 18 and 23 residues were present between 1-5%. Random clone picking and Sanger sequencing of 30 phagemids from round-3 and round-4 phage pools identified 9 sequence-unique Fab clones. These clones had CDRH3 lengths of 7, 8, 10, 14, 17 and 18 residues. We reconstructed 7 rare clones from five less-frequent CDRH3 lengths (11, 12, 15, 16, 18 and 23 residues). The frequencies of reconstructed rare clones ranged from 0.7% to 0.1%. Phage-ELISA indicated that 15 out of 16 clones bound to Jagged-2 but not to the control proteins **(Figure 5.12B)**. One rare clone (J2/F/R1) bound to BSA and the Fc protein in addition to Jagged-2. We converted 13 phage-displayed Fabs into soluble Fabs encompassing different CDRL3 and CDRH3 lengths, and assessed their affinity and specificity using Fab-ELISA **(Figure 5.12C)**. Among the seven top clones, five Fabs possessed low-nM EC_{50} values for Jagged-2. However, all five Fabs cross-reacted with Jagged-1. J2/F/5 and J2/F/8 had low affinities for Jagged-2, nonetheless had high Fab expression rates in *E. coli*. It is worth noting that J2/F/5 had the highest frequency in the round-3 phage pool (45%). Among the six rare clones, five Fabs possessed low-nM EC_{50} values for Jagged-2. The rare clone J2/F/R7 with the lowest frequency (0.1%) had an EC_{50} of 24.62 nM. Strikingly, three out of six rare clones (J2/F/R4, J2/F/R6 and J2/F/R7) possessed strict specificity for Jagged-2. Together, these results indicated that rare clones could not only bind tighter than top clones but also could be more specific than top clones.

5.1.8 Evolution Profiles of Top Clones and Rare Clones

To monitor the evolution profiles of Fabs during phage display selections, we sequenced the CDRH3 region of Notch/Jagged selection outputs on a 100 bp Ion semiconductor chip, and followed the frequency of CDRH3 sequences in selection outputs over five subsequent rounds of selection. CDRH3 sequences corresponding to 24 ELISA-positive top clones and 24 ELISA-positive rare clones were chosen for this analysis from both Fab libraries (S and F). The propagation behavior for 24 top sequences and 24 rare sequences are shown in **Figures 5.13A and 5.13B**. The average propagation behavior for 24 top sequences and 24 rare sequences are shown in **Figures 5.13C and 5.13D**. On average, top sequences showed significant enrichments in rounds- 3, 4 and 5 and reached 21.7% after five rounds. Rare sequences showed a minor enrichment in round-3 (1.4%) and depletion in rounds- 4 and 5. This confirmed that three selection rounds would be sufficient to isolate top clones and reconstruct rare clones from phage

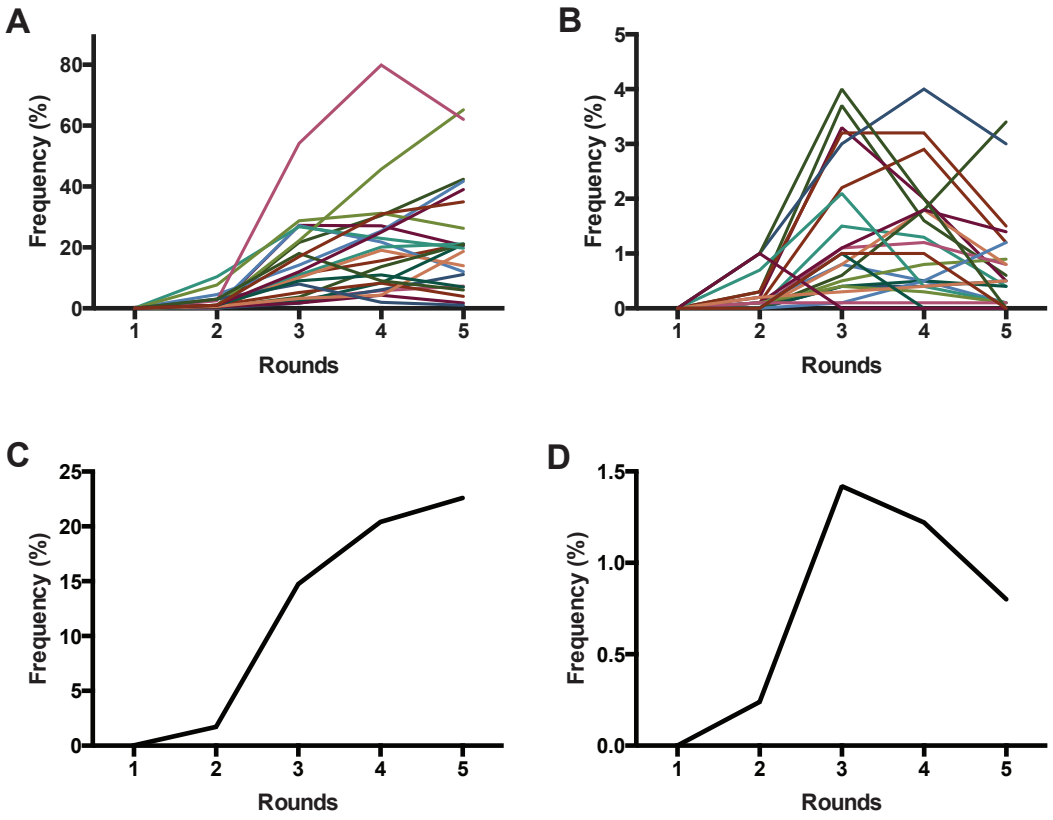


Figure 5.13: Propagation behavior of **(A)** 24 top CDRH3 sequences and **(B)** 24 rare CDRH3 sequences. X-axis indicates rounds 1 to 5, and Y-axis indicates the frequency of Fab CDRH3 sequences in selection outputs. Average propagation behavior of **(C)** 24 top CDRH3 sequences and **(D)** 24 rare CDRH3 sequences.

selection outputs. Rounds- 4 and 5 increased the abundance of top clones at the expense of rare clones. In addition to highlighting the differences in evolution profiles between top and rare clones, this analysis clearly illustrated the need for NGS-based approaches to recover rare clones from selection outputs.

5.1.9 Discussion

In recent years, NGS technologies have revolutionized both fundamental and translational aspects of biological research. Sequencing of functional antibody repertoires and subsequent bioinformatic analyses have provided unprecedented insight into the mechanisms of B-cell development in humans and other species, and has transformed our understanding of immune responses to autoimmunity, vaccination, infection and cancer. Sequencing of antibody repertoires is also used for identifying critical biomarkers and epitopes targeted by functional

antibody responses, for developing vaccines, immunomodulatory drugs and diagnostic tools, and for discovering antibodies (reviewed by Georgiou *et al*., 2014; Robinson, 2015; Lavinder *et al*., 2015). NGS technologies have also been employed in several phage display platforms for analyzing natural and synthetic antibody fragment sequences, and for identifying and reconstructing scFv and Fab binders not found by conventional ELISA screens (reviewed by Naso *et al*., 2014; Glanville *et al*., 2015).

In this work, we developed an NGS-assisted antibody discovery platform by integrating phage-displayed single-framework synthetic Fab libraries with Ion Torrent sequencing. The use of single-framework synthetic Fab libraries enabled seamless integration of antibody phage display with short-read DNA sequencing. With a defined single-framework, sequencing efforts were directed only towards the diversified CDR regions. Since NGS primers were designed to anneal to the framework regions, single-framework libraries also offered the advantage of having one set of primers during amplicon preparation. During sequence analysis, the use of a single-framework eliminated complications associated with the occurrence of multiple frameworks and framework mutations present in natural antibody repertoires. To compensate for the reduction in variable domain diversity, the single-framework libraries were rationally engineered to include length, amino acid and conformational diversities within CDRs. Using trinucleotide mutagenesis technology (Virnekas *et al*., 1994), two or four of the six CDR loops were diversified in a combinatorial fashion based on structural bioinformatics analysis of antigen-antibody structures. Besides contributing to library stability and function, precise engineering of CDR diversity enabled fast and accurate interpretation of NGS data sets. The new Fab library described in this work delivered promising Fabs against Jagged-1/2 and was moderately successful against Notch-2/3. The highest-affinity Fab bound to Jagged-2 with a K_D value of 80 ± 13 pM and showed >1000-fold specificity over Jagged-1.

Between the two commonly used short-read DNA sequencing platforms, Illumina and Ion Torrent, we chose to use the latter. Ion Torrent sequencing can deliver up to 5.5×10^6 reads with 100-400 bp read lengths in a short span of time (5 hours) at low cost ($749 per run). MiSeq from Illumina can provide up to 25×10^6 reads with 2 × 300 bp read lengths (paired-end reads), however it takes up to 55 hours and $1750 for a sequencing run (reviewed by Glanville *et al*., 2015). With low cost and fast turnaround times, Ion Torrent sequencing enabled routine analysis of phage display selection outputs. NGS information was used for (1) assessing the quality of

Fab libraries, (2) analyzing selection outputs in terms of sequence evolution, length and amino acid distribution, and (3) identifying and reconstructing low-frequency rare clones from selection outputs.

Due to limitations in attainable read lengths, NGS of Fab libraries and selection outputs is usually restricted to CDRH3 (Ravn *et al*., 2010; D'Angelo *et al*., 2014). Since CDRH3 sequencing alone is not sufficient for high-throughput reconstruction of Fabs, we developed a method for linking and sequencing all diversified CDRs in phage Fab pools without losing right CDR combinations. Our method determined the sequences of paired CDRs in selection outputs and paved the way for straightforward conversion of NGS information into Fab clones. The method is rapid and simple and can be adapted to any synthetic scaffold-based phage library. Previous studies have shown that single-framework synthetic Fab libraries allow facile reformatting of Fabs between different vector systems for affinity maturation, bacterial expression and IgG conversion (Miersch and Sidhu, 2012; Adams and Sidhu, 2014). In this work, we demonstrate an additional advantage of single-framework synthetic Fab libraries. The combination of single Fab framework and precise CDR diversity resulted in rapid and reliable reconstruction of Fabs from NGS information.

Typically, to recover antigen-specific Fabs from phage selection pools, hundreds to thousands of colonies from selection outcomes are first interrogated by binding assays (phage/Fab ELISA) in 96/384 well plates, and then identities of ELISA-positive clones are determined by Sanger sequencing (Ravn *et al*., 2010). Due to phage propagation biases in *E. coli*, many binders present in earlier selection rounds do not get enriched and remain at very low frequencies in later selection rounds. Therefore, the conventional clone recovery method leads to the repeated identification of the same enriched clones from phage selection pools (Derda *et al*., 2011; Saggy *et al*., 2012). NGS-based reconstruction of antibody clones can avoid the repeated identification of growth-advantaged, high frequency binders, and can identify and recover many unique clones that are present in low frequencies in phage pools (Ravn *et al*., 2013; Naso *et al*., 2014; Spiliotopoulos *et al*., 2015; Lopez *et al*., 2017). As evidence of this, low-frequency rare Fabs (~0.1%) reconstructed in this work showed higher affinity and specificity than high-frequency top Fabs isolated by Sanger sequencing. Reconstruction of rare Fab clones from less-frequent CDRH3 lengths increased the possibility of recovering Fabs with new and better binding solutions.

Synthetic antibody technology is well known for high-throughput generation of antibodies. In this work, the following factors contributed to the high-throughput reconstruction and Fabs from selection outputs: (1) multiplexing of up to 20 selection outputs in one sequencing run; (2) semi-automated NGS data processing and analysis; (3) synthesis of CDR-encoding oligonucleotides in 96-well format; (4) optimized procedures for cloning CDR-encoding oligonucleotides into the Fab-encoding template phagemid; (5) replacement of Sanger sequencing with restriction digestion analysis for screening positive phagemid clones; (6) optimized procedures for sub-cloning ELISA-positive phage-Fab clones into Fab-expression vector in 96-well format and (7) confirmation of Fab sub-cloning and expression by bio-layer interferometry in 96-well format. We isolated top Fab clones by Sanger sequencing and compared them with rare Fab clones for emphasizing the need for NGS-based reconstruction approaches in antibody phage display. This step is now excluded in routine phage display selections. Following three rounds of selection, the selection output is subjected to L3-H3 or L3-H1-H2-H3 sequencing, promising clones are identified by NGS analysis, reconstructed and tested for binding. This approach eliminated the use of Sanger sequencing from antibody phage display (from phage selections to Fab K_D determination) and resulted in a reliable and high-throughput antibody discovery platform.

5.2 Systematic Generation of Synthetic Antibody Fragments with Optimal CDR Lengths for Notch-1 Recognition

In section 5.1, we validated the functionality of library-S by generating high-affinity Fabs against Notch and Jagged receptors (Notch-2/3 and Jagged-1/2). The highest-affinity Fab bound to Jagged-2 with a K_D value of 80 ± 13 pM and showed >1000-fold specificity over Jagged-1. Though Notch-2/3 Fabs had low-nM K_D values, they showed cross-reactivity with other Notch and Jagged members. In this section, we describe the generation and validation of anti-Notch-1 Fabs from two phage-displayed single-framework synthetic Fab libraries, library-S and library-F. Over the course of Fab generation, we also show that implementing NGS approaches, screening focused diversity libraries, fine-tuning the library diversity and making small changes to diversity designs can improve the success rate of single-framework synthetic Fab libraries.

5.2.1 Notch-1 Fabs from Library-S

Library-S contains four fixed CDRs and two diversified CDRs (CDRL3 and CDRH3) on an optimized Hu4D5-8 framework. Length diversity was included within CDRH3 and amino acid diversity was included within both the diversified CDRs **(Figures 5.14A and 5.14B)**. To generate Notch-1 Fabs, we conducted solid-phase panning of library-S against the human Notch-1 receptor extracellular domain. Briefly, we immobilized the recombinant Notch-1-Fc fusion protein on Maxisorp plates, and incubated target-coated wells with library-S deselected for binding to the Fc protein. Upon elimination of non-specific phages by washing, we eluted bound phages and amplified them in *E. coli* overnight for subsequent rounds of panning. Random clone picking and Sanger sequencing of 20 phagemids from the round-4 phage pool gave rise to two Fab clones **(Figure 5.14C)**. Clones N1/S/1 and N1/S/2 had a CDRH3 length of 13 and 16 residues, respectively. Phage-ELISA indicated that both the Fabs bound to Notch-1 but not to BSA and the Fc protein (data not shown).

Since lengths of five other CDRs were fixed, we anticipated that CDRH3 length diversity would play a central role in library-S function. To obtain Notch-1 Fabs with different CDRH3 lengths, we split library-S into four sub-libraries each containing a set of CDRH3 lengths, and panned them against Notch-1. After each selection round, we calculated the number of phages eluted from target-coated wells relative to BSA-coated wells. A positive enrichment in target-specific phage number was only observed with two sub-libraries SL2 and SL3 **(Figure 5.14D)**.

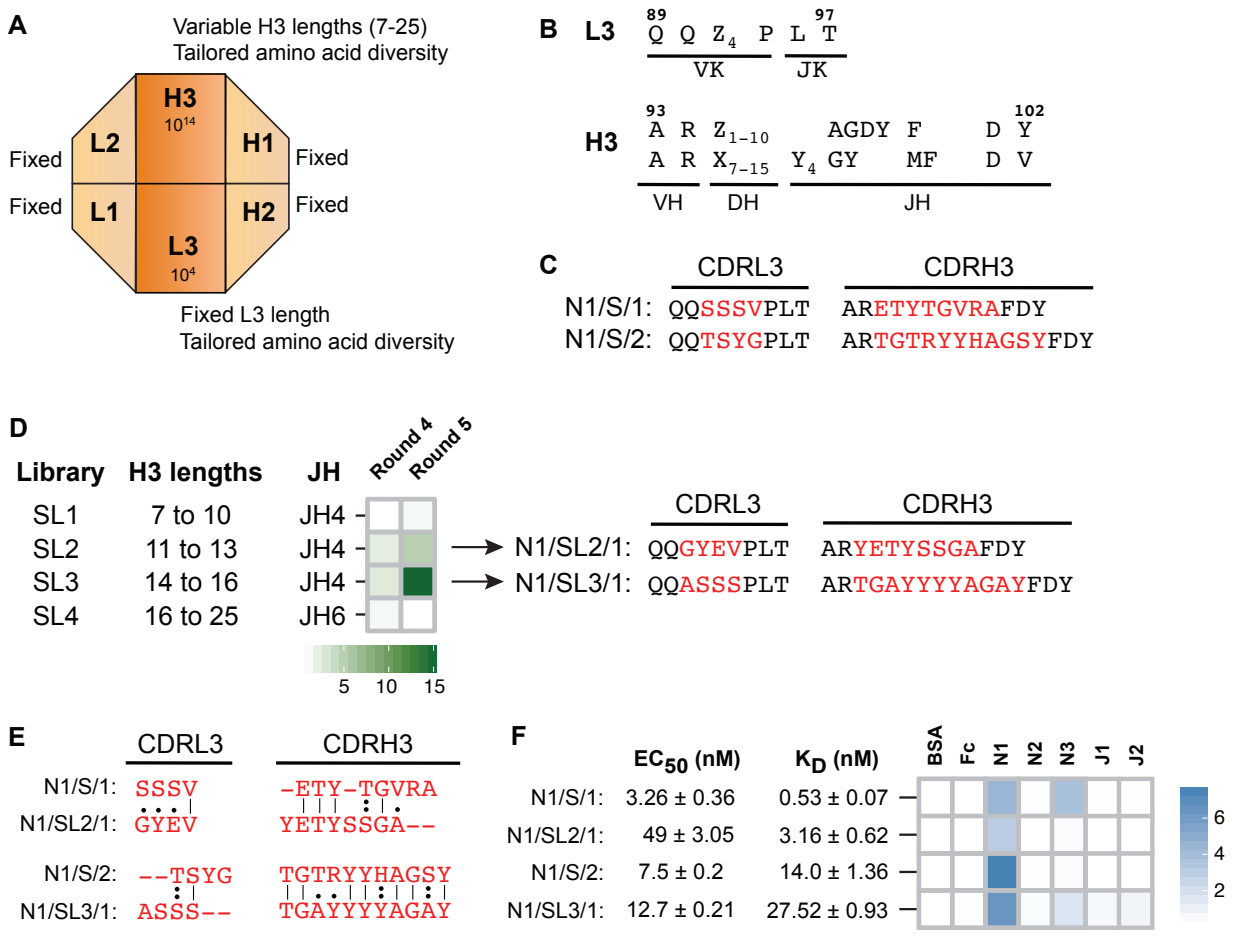


Figure 5.14: Notch-1 Fabs from library-S. (A) Schematic representation of library-S CDR diversity. **(B)** CDR diversity designs for L3 and H3. Z denotes any of the following thirteen amino acids introduced at different proportions: Y (20%), S (20%), G (20%), T (6.5%), A (6.5%), P (6.5%), H (3.5%), R (3.5%), E (3.5%), F (2.5%), W (2.5%), V (2.5%) or L (2.5%). X denotes any of the following nine amino acids introduced at different proportions: Y (25%), S (20%), G (20%), A (10%), F (5%), W (5%), H (5%), P (5%) or V (5%). CDRH3 length is varied by altering the number of Z and X. **(C)** CDRL3 and CDRH3 sequences of two Notch-1 Fabs isolated from library-S. Fixed anchor residues are in black and diversified CDR positions are in red. **(D)** Left panel: panning of library-S sub-libraries (SL1 to SL4) against Notch-1. Heatmap showing the enrichment of Notch-1 binding phages after phage display selection rounds 4 and 5. Fold enrichment is the ratio of number of phages eluted from target-coated wells to number of phages eluted from BSA-coated wells. Right panel: CDRL3 and CDRH3 sequences of Fab N1/SL2/1 (isolated from SL2) and Fab N1/SL3/1 (isolated from SL3). **(E)** Pairwise sequence alignments between master- and sub-library Fab CDR sequences. Only diversified CDR positions within L3 and H3 were included for the alignment. Sequences were aligned in EMBOSS Needle using the Needleman-Wunsch algorithm. **(F)** Affinity and specificity of four Notch-1 Fabs isolated from library-S master- and sub-libraries. EC_{50} and K_D values were determined using multi-point Fab-ELISA and bio-layer interferometry, respectively. Fab specificity was determined using single-point Fab-ELISA at 1 μM Fab concentration. ABS_{450} values are shown as a heatmap.

Interestingly, SL2 and SL3 contained CDRH3 lengths of 13 and 16 amino acids, respectively, within their length set. Through random clone picking and Sanger sequencing, we isolated Fab clones N1/SL2/1 and N1/SL3/1 from SL2 and SL3 round-4 phage selection pools, respectively **(Figure 5.14D)**. N1/SL2/1 and N1/S/1 had the same CDRH3 length of 13 residues. Their CDRL3 central segments were 25% similar and their CDRH3 diversity segments were 50% similar to each other. Both N1/SL2/1 and N1/S/1 contained the ETY-sequence at the N-terminus of CDRH3. N1/SL3/1 and N1/S/2 had the same CDRH3 length of 16 residues. Their CDRL3 central segments were 33.3% similar and their CDRH3 diversity segments were 81.8% similar to each other. N1/SL3/1 contained SSS-sequence within its CDRL3, which was also found in N1/S/1 **(Figure 5.14E)**. Phage-ELISA indicated that N1/SL2/1 and N1/SL3/1 bound to Notch-1 but not to BSA and the Fc protein (data not shown). Random clone picking and Sanger sequencing of 20 phagemids from SL1 and SL4 round-4 selection pools gave rise to six Fab clones, however they bound to all test and control antigens in phage-ELISA (data not shown).

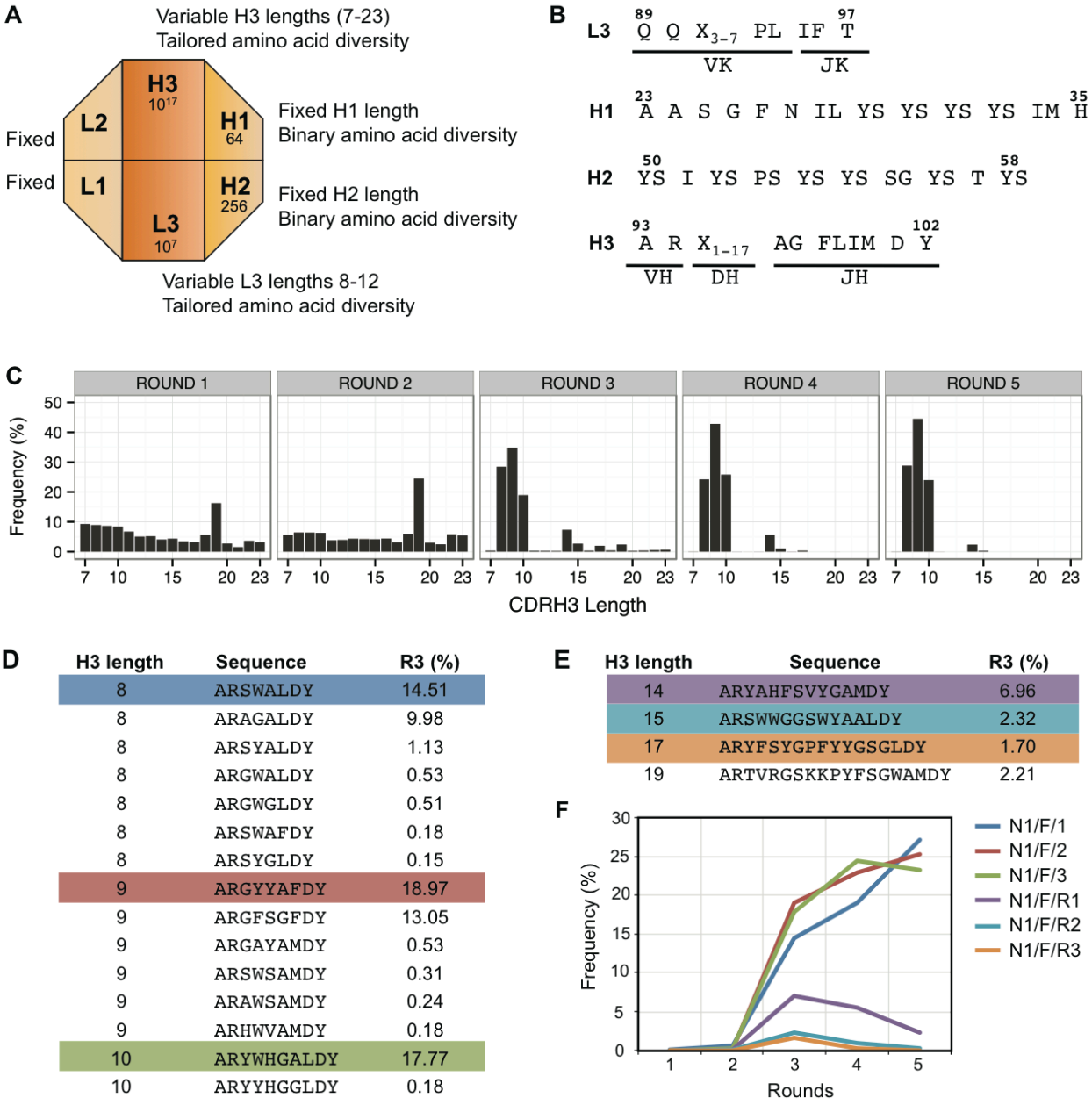
To determine Fab affinity and specificity, we subcloned the four Notch-1 phage clones into a Fab expression vector and expressed and purified them from *E. coli* using Protein-L affinity chromatography. First, we determined EC_{50} values for Fabs binding to Notch-1 using Fab-ELISA. The EC_{50} values for N1/S/1 and N1/S/2 were 3.3 nM and 7.5 nM, respectively. With an EC_{50} of 49 nM, N1/SL2/1 bound 15-fold weaker than N1/S/1. The EC_{50} of N1/SL3/1 (12.7 nM) was similar to that of N1/S/2 **(Figure 5.14F)**. Second, we measured the kinetics of Fab binding using bio-layer interferometry. Briefly, Fabs were immobilized on amine-reactive sensors and sensor tips were exposed to increasing concentrations of Notch-1. Association and dissociation rates were assessed by a wavelength shift, and kinetic data sets were fit using a 1:1 binding model. Consistent with Fab-ELISA, N1/S/1 was the highest-affinity binder (K_D = 0.53 nM). N1/SL2/1 bound to Notch-1 with a K_D value of 3.2 nM. N1/S/2 and N1/SL3/1 possessed mid-nM K_D values, 14 nM and 27.5 nM, respectively **(Figure 5.14F)**. In both cases, the sub-library Fabs showed weaker binding than the corresponding master-library Fabs. Third, to assess Fab specificity, we tested the binding of Fabs to Notch and Jagged receptor ectodomains using Fab-ELISA. At 1 μM Fab concentration, N1/S/1 and N1/SL3/1 cross-reacted with Notch-3, whereas N1/SL2/1 and N1/S/2 exhibited target-specific binding **(Figure 5.14F)**. Kinetic analysis indicated that the highest-affinity Fab N1/S/1 bound to Notch-3 with a K_D of 14.1 ± 0.53 nM. Interestingly, the specificity of N1/SL2/1 was better than its corresponding master-library Fab

N1/S/1, and the specificity of N1/S/2 was better than its corresponding sub-library Fab N1/SL3/1.

5.2.2 Notch-1 Fabs from Library-F

To obtain Notch-1 Fabs with both sub-nM affinity and high specificity, we decided to use library-F for selections. Library-F contains four diversified CDRs and two fixed CDRs on an optimized Hu4D5-8 framework. Length diversity was included within CDRs L3 and H3, and amino acid diversity was included within CDRs L3 and H3 and solvent-accessible positions of CDRs H1 and H2 (Persson *et al*., 2013) **(Figures 5.15A and 5.15B)**. We anticipated that the additional diversity in library-F (length diversity in CDRL3 and amino acid diversity in CDRs H1/H2) would generate more potent and selective Fabs for Notch-1.

Solid-phase panning of library-F was conducted against the Notch-1 ectodomain. In an effort to identify Fabs with different CDRH3 lengths, we subjected the CDRH3 region of the phage selection outputs to Ion Torrent sequencing. Briefly, we amplified CDRH3 from the phage pools, performed emulsion amplification of the amplicon library on Ion sphere particles (ISPs), and sequenced the enriched ISPs on a 100 bp Ion semiconductor chip. First, we monitored the changes in CDRH3 length distribution over five subsequent rounds of selection **(Figure 5.15C)**. With a clear enrichment in three short CDRH3 lengths (8, 9 and 10 residues), the length distribution changed significantly after three rounds. Also, four other CDRH3 lengths (14, 15, 17 and 19 residues) stood out from the rest of the population with frequencies ranging between 2-7%. Next, we ranked round-3 CDRH3 sequences based on their relative frequencies. 19 sequences were present above 0.1%, and all of them had the above-mentioned CDRH3 lengths. While 15 out of 19 sequences had short CDRH3 lengths **(Figure 5.15D)**, each of the other four CDRH3 lengths had one sequence above 0.1% **(Figure 5.15E)**. The longest one in the list with 19 residues turned out to be the anti-maltose binding protein CDRH3 sequence used in the template phagemid. Next, we decided to isolate the most-frequent Fab from each selected CDRH3 length, and followed the frequency of the six chosen Fab CDRH3 sequences over five subsequent rounds of selection **(Figure 5.15F)**. The three short CDRH3 sequences showed a significant enrichment throughout the selection process. The three medium-sized CDRH3 sequences not only showed a lower enrichment in round-3 but also showed depletion in the later rounds.



H3 length	Sequence	R3 (%)
8	ARSWALDY	14.51
8	ARAGALDY	9.98
8	ARSYALDY	1.13
8	ARGWALDY	0.53
8	ARGWGLDY	0.51
8	ARSWAFDY	0.18
8	ARSYGLDY	0.15
9	ARGYYAFDY	18.97
9	ARGFSGFDY	13.05
9	ARGAYAMDY	0.53
9	ARSWSAMDY	0.31
9	ARAWSAMDY	0.24
9	ARHWVAMDY	0.18
10	ARYWHGALDY	17.77
10	ARYYHGGLDY	0.18

H3 length	Sequence	R3 (%)
14	ARYAHFSVYGAMDY	6.96
15	ARSWWGGSWYAALDY	2.32
17	ARYFSYGPFYYGSGLDY	1.70
19	ARTVRGSKKPYFSGWAMDY	2.21

Figure 5.15: Panning library-F against Notch-1. (A) Schematic representation of library-F CDR diversity. **(B)** CDR diversity designs for L3, H1, H2 and H3. X denotes any of the following nine amino acids introduced at different proportions: Y (25%), S (20%), G (20%), A (10%), F (5%), W (5%), H (5%), P (5%) or V (5%). CDRL3 and CDRH3 lengths are varied by altering the number of X. **(C)** Changes in CDRH3 length distribution over five subsequent phage display rounds for the Notch-1 selection. The peak in rounds 1 and 2 (CDRH3 length of 19 residues) is due to the presence of anti-MBP Fab CDRH3 sequence from the template phagemid. **(D)** CDRH3 sequences >0.1% in round-3 with short loop lengths (8, 9 and 10 amino acids). **(E)** CDRH3 sequences >0.1% in round-3 with loop lengths of 14, 15, 17 and 19 amino acids. In panels D and E, the most abundant sequence from each CDRH3 length is highlighted. **(F)** Propagation behavior of highlighted CDRH3 sequences over five rounds of selection.

As expected, random clone picking and Sanger sequencing of 20 clones from round-3 and round-4 phage pools only recovered the three most-frequent clones with short CDRH3 lengths (referred to as top clones). Therefore, we used our NGS-assisted Fab reconstruction method to recover the three less-frequent clones with medium CDRH3 lengths (referred to as rare clones). First, to identify the diversified CDR sequences (L3, H1 and H2) that paired with the CDRH3s of interest, we linked the diversified CDRs in the round-3 phage pool DNA next to each other by deleting the intervening framework regions, and sequenced all four diversified CDRs as a CDR strip (L3-H1-H2-H3) on a 400 bp Ion semiconductor chip. As the maximum read-length offered by Ion Torrent is only 400 bases, the framework deletion step was necessary to reduce the amplicon length from 1100 bases to 200 bases. Once the CDR-combinations were identified, we reconstructed the desired Fab clones by cloning CDR-encoding oligonucleotides into a Fab-4D5 encoding template phagemid by Kunkel mutagenesis. The diversified CDR sequences of all the six Fab clones are shown in **Figure 5.16A**.

In phage-ELISA, all the six Fab clones tested positive for binding to Notch-1 (**Figure 5.16A)**. Compared to the top clones, ELISA signals were ~4-fold weaker for N1/F/R2, and ~10-fold weaker for N1/F/R3. Next, we converted the six phage-Fab clones into soluble Fab proteins, and measured EC_{50} values for Fabs binding to Notch-1 using Fab-ELISA (**Figure 5.16B)**. Except for N1/F/R3, the other five Fabs had an EC_{50} of ~4 nM. N1/F/R3 bound ~10-fold weaker than the other Fabs (EC_{50} = 37 nM). Next, we analyzed Fab-binding kinetics using bio-layer interferometry and Fab specificity using Fab-ELISA (**Figure 5.16B)**. Even though the two top clones N1/F/1 and N1/F/2 exhibited sub-nM K_D values for Notch-1, they also possessed mid-nM K_D values for Notch-2, and low-nM values for Notch-3. N1/F/1 bound to Notch-1, Notch-2 and Notch-3 with K_D values of 0.45 ± 0.05 nM, 40.68 ± 1.24 nM and 5.3 ± 0.22 nM, respectively. N1/F/1 affinity was 90-fold lower for Notch-2 and 12-fold lower for Notch-3. N1/F/2 bound to Notch-1, Notch-2 and Notch-3 with K_D values of 0.72 ± 0.05 nM, 30.92 ± 0.96 nM and 4.31 ± 0.17 nM, respectively. N1/F/2 affinity was 43-fold lower for Notch-2 and 6-fold lower for Notch-3. In brief, the specificity of N1/F/1 was two-fold better than N1/F/2. The third top clone N1/F/3 had a K_D value of 1.88 nM for Notch-1, and cross-reacted with Notch-3 and Jagged-2 in Fab-ELISA. The three rare clones with medium CDRH3 lengths possessed very similar binding profiles. They bound to Notch-1 with low-nM K_D values, and showed weak to moderate binding with Jagged-2 in Fab-ELISA. N1/F/R1 and N1/F/R2 bound to Notch-1 with K_D values of 1.98

nM and 1.42 nM, respectively. In agreement with phage-ELISA and Fab-ELISA, N1/F/R3 was the lowest-affinity Notch-1 binder from library-F (K_D = 6.6 nM). Kinetic analysis confirmed dose-dependent binding of N1/F/R2 with Jagged-2, and determined the K_D value to be 71.72 ± 2.18 nM.

Most notably, the rare clones recovered using NGS approaches had low-nM affinities for Notch-1, and were more specific than the top clones isolated by Sanger sequencing. In particular, N1/F/R2 was less frequent in the round-3 phage pool (~2%), did not show enrichment in round-4 and round-5, and had weak phage-ELISA output, however the K_D value for purified Fab N1/F/R2 was very similar to N1/F/3 and N1/F/R1 that were more frequent in phage pools. In addition, N1/F/R2 showed tighter binding to Jagged-2 than other clones. This example clearly illustrates the need for substituting ELISA-based screening with NGS-based strategies for identifying Fab leads from phage selection outputs.

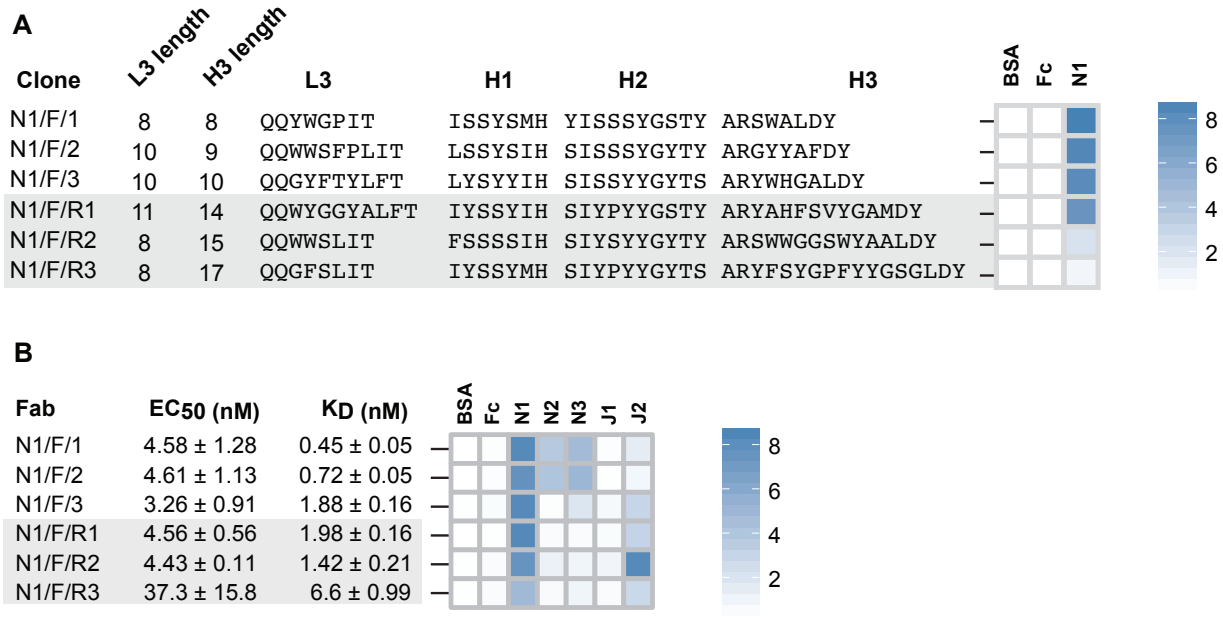


Figure 5.16: Notch-1 Fabs from library-F. (A) Sequence characteristics and phage-displayed Fab binding characteristics of 3 top Fabs isolated by Sanger sequencing and 3 rare Fabs reconstructed from L3-H1-H2-H3 NGS information. **(B)** Affinity and specificity of purified Notch-1 Fabs. Fab specificity was determined at 1 μM Fab concentration. Phage-ELISA and specificity-ELISA ABS_{450} values are shown as heatmaps. Rare clones are highlighted in gray.

5.2.3 Notch-1 Fabs from the Modified-F Long-CDR Library

In addition to facilitating rare clone reconstruction, sequencing CDR-combinations in a selection output opens the possibility of studying how CDRs are paired between each other against a given target. Here, to check the influence of CDRL3 length diversity on CDRH3 length solutions for Notch-1, we monitored the pairing between CDRL3 lengths and CDRH3 lengths in the round-3 phage selection output. The analysis showed a clear indication for a preferential pairing of certain CDRL3 and CDRH3 lengths **(Figure 5.17A)**. For example, CDRH3 length of 8 amino acids most-frequently paired with CDRL3 lengths of 8 and 9 amino acids. Also, it appeared that the selection preferred to use the longest CDRL3 length to provide long CDRH3 length solutions for Notch-1 recognition. Since Fabs with long CDRH3 sequences occurred at very low frequencies (copy number <3), we did not reconstruct these clones. In addition to slower amplification rates in *E. coli*, the poor enrichment of long CDR clones could result from the underrepresentation of long CDR sequences in the library. To confirm this, we amplified the CDRL3 and CDRH3 regions from the naïve library-F phage pool, and sequenced them on a 100 bp Ion semiconductor chip. Length analysis confirmed the presence of a bias in the library towards short sequences in both the CDRs **(Figure 5.17B)**. Also, we noticed high retention of template CDRL3 (~24%) and CDRH3 (~13%) sequences in the library (bars corresponding to CDRL3 length of 9 residues and CDRH3 length of 19 residues in Figure 5.17B).

To isolate Notch-1 Fabs with long CDRs, we decided to modify the length bias in library-F towards long CDRs, and use the modified library for selections. The modified-F library had the following design features: (1) CDRL3, CDRH1 and CDRH2 diversity designs were unaltered; (2) in addition to 17 JH4-CDRH3 lengths (7-23 residues) used in library-F, 10 JH6-CDRH3 lengths (16-25 residues) used in library-S were included; (3) to bias the length diversity towards long CDR lengths, the amount of mutagenic oligonucleotides used in mutagenesis reactions were normalized according to their theoretical diversity; and (4) to prevent the expression of Fabs with unmutated template CDRs, the anti-maltose-binding protein diversified CDR sequences were replaced with TAA stop codons in the template phagemid. The modified-F library was constructed using a well-established M13 bacteriophage system that allows bivalent Fab display (Lee *et al*., 2004A). Kunkel mutagenesis was used to repair the stop codons in the Fab-4D5 encoding template phagemid and replace CDR positions with fixed or degenerate codons encoding the designed amino acid composition. Following mutagenesis, the library DNA was

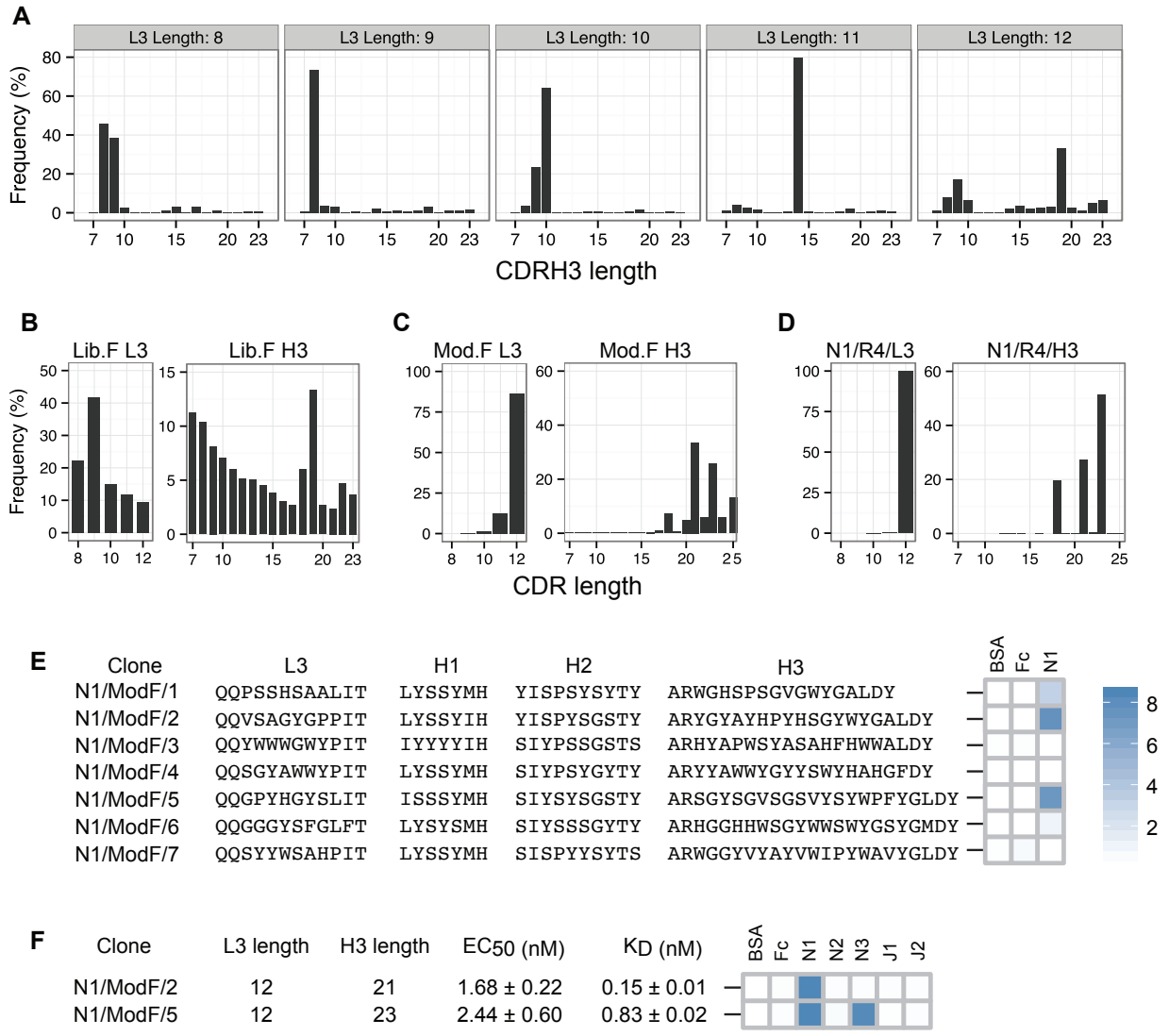


Figure 5.17: Notch-1 Fabs from the modified-F long-CDR library. (A) Pairing between CDRL3 lengths and CDRH3 lengths in the Notch-1 selection output (round-3) from library-F selections. L3-H3 pairs were obtained from L3-H1-H2-H3 NGS information, and for each possible L3 length, H3 length distribution was generated. **(B)** CDRL3 and CDRH3 length distribution of the naïve library-F. **(C)** CDRL3 and CDRH3 length distribution of the modified-F library. **(D)** CDRL3 and CDRH3 length distribution of the Notch-1 selection output after four rounds of selection with the modified-F library. **(E)** Diversified CDR sequences and phage-displayed Fab binding characteristics of 7 Notch-1 Fabs isolated from modified-F selections using Sanger sequencing. **(F)** CDRL3 length, CDRH3 length, affinity and specificity of Fabs N1/ModF/2 and N1/ModF/5. EC_{50} and K_D values were determined by multi-point Fab-ELISA and bio-layer interferometry, respectively. Fab specificity was determined at 1 μM Fab concentration. Phage-ELISA and specificity-ELISA ABS_{450} values are shown as heatmaps.

electroporated into an *E. coli* strain suitable for high-efficiency transformation and phage production. NGS analysis of the CDRL3 and CDRH3 regions of the naïve modified-F library confirmed proper CDR diversification and the length bias towards long CDRs **(Figure 5.17C)**.

We panned the modified-F long-CDR library against Notch-1, and sequenced the CDRL3 and CDRH3 regions from the round-4 phage selection output on a 100 bp Ion semiconductor chip. Length analysis indicated a significant enrichment in CDRL3 sequences of 12 residues and CDRH3 sequences of 18, 21 and 23 residues **(Figure 5.17D)**. Random clone picking and Sanger sequencing of 20 phagemids recovered seven unique clones from the round-4 phage pool. Three clones tested positive for binding to Notch-1 in phage-ELISA **(Figure 5.17E)**. Each ELISA-positive clone had a different CDRH3 length, and also showed a high CDRH3 sequence enrichment in the round-4 phage pool (19.15% for N1/ModF/1, 25.96% for N1/ModF/2, and 43.5% for N1/ModF/5). We chose to pursue with two clones that had longer CDRH3 lengths (N1/ModF/2 and N1/ModF/5). We converted the phage-Fab clones into soluble Fab proteins, and assayed them using Fab-ELISA **(Figure 5.17F)**. Both the Fabs bound stronger to Notch-1 than all other library-S or library-F Fabs. The EC_{50} values for N1/ModF/2 and N1/ModF/5 were 1.7 nM and 2.4 nM, respectively. Further, N1/ModF/2 showed strict specificity for Notch-1, whereas N1/ModF/5 showed dual specificity for Notch-1 and Notch-3. During kinetic studies, N1/ModF/2 exhibited the highest affinity for Notch-1 (K_D = 0.15 nM). N1/ModF/2 possessed ~17-fold higher affinity than the Notch-1 specific antibody anti-NRR1 reported previously (K_D = 2.5 nM) (Wu *et al*., 2010). With a sub-nM K_D and strict specificity, N1/ModF/2 turned out to be the best Fab clone for Notch-1. N1/ModF/5 bound to Notch-1 and Notch-3 with K_D values of 0.83 nM and 7.1 nM, respectively.

5.2.4 MERTK Fab from the Modified-F Long-CDR Library

To assess the performance of modified-F against Notch-unrelated targets, we conducted solid-phase selections against four cell-surface receptor ectodomains (GP130, VEGFR-1, IL6Rα, and MERTK). After each selection round, we calculated the number of phages eluted from target-coated wells relative to BSA-coated wells **(Figure 5.18A)**. A positive enrichment in target-specific phages was not observed for three selections. This could have been due to poor target behavior on Maxisorp plates, absence of right library diversity and/or biases in phage propagation during selection rounds. Further studies are required to draw conclusions about these selection results. A moderate enrichment was observed against MERTK, a target for which both

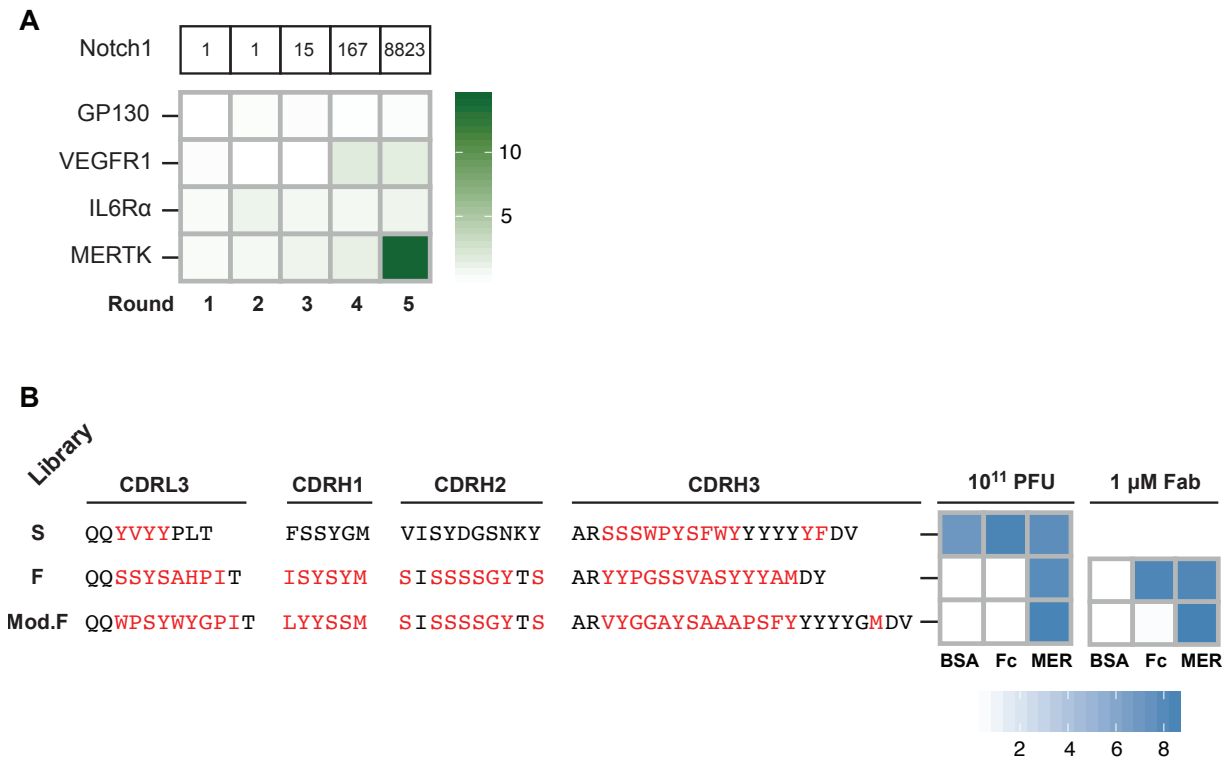


Figure 5.18: Performance of the modified-F long-CDR library against Notch-unrelated targets. (A) Heatmap showing the enrichment of target-binding phages for four targets over five selection rounds. Enrichment values from the Notch-1 selection are given for comparison. Fold enrichment is the ratio of number of phages eluted from target-coated wells to number of phages eluted from BSA-coated wells. **(B)** CDR sequences, phage-displayed Fab and purified-Fab binding profiles of MERTK Fabs isolated from library-S, library-F and the modified-F library. Within CDR sequences, fixed positions are in black and diversified positions are in red. Phage-ELISA and Fab-ELISA ABS_{450} values are shown as heatmaps.

the other libraries (S and F) proved to be unsuccessful. Through random clone picking and Sanger sequencing, we isolated a MERTK Fab from the modified-F selection, which bound to MERTK with an EC_{50} of 4.86 ± 0.85 nM and a K_D of 1.68 ± 0.22 nM. Interestingly, the clone contained the design features found only in modified-F, but not in library-S (CDRL3 length of 12 residues and H1/H2 optimization) or library-F (CDRH3 length of 24 residues and JH6 usage) **(Figure 5.18B)**. The results from MERTK selections indicated that small changes to diversity designs could increase the success rate of single-framework synthetic Fab libraries. Overall, the modified-F library might not be suitable against all kinds of targets, however could be used whenever there is a need for long CDR length solutions or upon failure with other libraries.

5.2.5 Discussion

Given that library-S delivered Fabs against Jagged-1/2 and Notch-2/3, we tested whether library-S could give rise to Fabs with high affinity and specificity against Notch-1. Library-S gave rise to a sub-nM bi-specific binder (N1/S/1) and a mid-nM mono-specific binder (N1/S/2). Next, we split library-S into four sub-libraries each containing a set of CDRH3 lengths, and panned them against Notch-1. Through sub-library screens, we expected to obtain Fabs with different CDRH3 lengths and Fabs with higher affinities compared to the master-library Fabs. Although results were contrary to our expectations, sub-library panning gave rise to a low-nM mono-specific binder (N1/SL2/1). Its corresponding master-library Fab showed six-fold higher affinity for Notch-1 (sub-nM K_D), however also had a mid-nM K_D for Notch-3. In going from library-S to library-F, we anticipated that the additional diversity in library-F (length diversity in CDRL3 and amino acid diversity in CDRs H1/H2) would generate more potent and selective Fabs for Notch-1. Through library-F selections, we isolated three top clones with short CDRH3 lengths by Sanger sequencing, and reconstructed three rare clones with medium CDRH3 lengths from NGS information. The top clones possessed sub-nM to low-nM K_D values for Notch-1 but also cross-reacted with two other receptors making them tri-specific binders. The rare clones possessed low-nM to mid-nM K_D values for Notch-1, and cross-reacted with Jagged-2 making them bi-specific binders.

Even though CDRH3 in library-S had higher length, amino acid and JH segment diversities than in library-F, library-S selections provided only two optimal CDRH3 length solutions for Notch-1. In library-F, the addition of length and conformational diversity to CDRs adjacent to CDRH3 increased the number of optimal CDRH3 length solutions for Notch-1. Further, CDRH3 length solutions allowed by library-F Fabs were distinct from library-S Fabs, indicating that length/conformational diversity in CDRs L3, H1 and H2 could also alter and determine the type of CDRH3 loop responsible for the interaction. Not only CDR lengths and amino acids, but also binding profiles of library-F Fabs were different from library-S Fabs, confirming previous observations that the increase in overall interface diversity fundamentally changes the nature of Fab-binding solutions rather than optimizing a common binding solution (Fellouse *et al*., 2007).

Since CDR length diversity in library-F was biased towards short sequences, we designed and constructed the modified-F library with the length bias towards long CDRs. We used the

modified-F library to isolate two Notch-1 Fabs (N1/ModF/2 and N1/ModF/5) with long CDRL3 and CDRH3 sequences. N1/ModF/2 exhibited the highest affinity for Notch-1 (K_D = 0.15 nM) and showed strict specificity for Notch-1. N1/ModF/5 showed low-nM binding to Notch-1 and cross-reacted with Notch-3. Strikingly, the binding profiles of modified-F Fabs were very similar to library-S Fabs. Only mono-specific Fabs (Notch-1) and bi-specific Fabs (Notch-1 and Notch-3) were isolated from both the libraries. N1/ModF/2 bound ~20-fold stronger than the best library-S-based mono-specific Fab N1/SL2/1. With a slightly weaker affinity for Notch-1 and two-fold higher affinity for Notch-3, N1/ModF/5 performed poorly than the library-S-based bi-specific Fab N1/S/1. While long CDR lengths contributed to the specificity of modified-F Fabs, we speculate that the additional amino acid diversity contributed to the specificity of library-S Fabs.

In this work, both single-framework synthetic Fab libraries (S and F) yielded sub-nM to mid-nM Notch-1 binders with different specificity profiles. However, to obtain high-affinity Notch-1 specific Fabs, screening the master library alone was not sufficient. In the case of library-S, screening sub-libraries identified a low-nM Notch-1 specific binder. In the case of library-F, fine-tuning the library towards long CDR lengths gave rise to a sub-nM Notch-1 specific binder. In addition to delivering Notch-1 Fabs, our work highlighted the importance of sampling a focused diversity within an antibody library. The master library only has sparse coverage of a large diversity space, due to limitations in attainable library size. Its sub-libraries provide a dense coverage of a smaller diversity space. If the master library is a lead discovery library, then its sub-libraries are targeted discovery libraries. Our work shows that random or rational focusing of a smaller diversity using targeted discovery libraries offer new and better binding solutions. In contrast, affinity maturation libraries offer optimization of an existing binding solution and are constructed by random or rational diversification of an existing antibody (Marvin and Lowman, 2015).

Previous studies have shown that NGS has three main applications in antibody phage display: library quality control, analysis of selection outputs, and reconstruction of rare clones (Ravn *et al*., 2010; Zhang *et al*., 2011; Mahon *et al*., 2013; D'Angelo *et al*., 2014; Glanville *et al*., 2015; Yang *et al*., 2017). In this work, NGS proved to be useful for all three applications. Most notably, we used NGS information for identifying and reconstructing rare clones with less-frequent CDRH3 lengths, for studying the pairing between CDRL3 and CDRH3 lengths in the

selection output, and for fine-tuning the library length diversity. These advances in the use of NGS information gave rise to Fab clones with better binding properties. We used CDRH3 length as a parameter for designing and screening libraries and for isolating Fabs from selection outputs. This approach proved to be successful for obtaining Fabs with different and superior binding profiles.

In summary, we used the synthetic antibody technology for generating potent and selective Fabs against the extracellular domain of Notch-1. Twelve Fabs with ten different CDRH3 lengths were identified from single-framework synthetic Fab libraries using phage display. Upon testing, Fabs showed high affinity for Notch receptors (sub-nM to mid-nM K_D values) and exhibited different binding profiles (mono- or bi- or tri-specific). Most likely, these Fabs recognize different epitopes on Notch-1 and could be used for modulating the Notch signaling pathway using different mechanisms of action. Two Fabs exhibited strict specificity for Notch-1 with clinically relevant K_D values. In contrast to gene knockout approaches, γ-secretase inhibitors and pan-Notch antibodies (Espinoza and Miele, 2013; Takebe *et al*., 2014), our mono-specific Fabs will permit more precise control of Notch-1 inhibition for both research and therapeutic applications. Over the course of Fab generation, we also showed that implementing NGS approaches, screening focused diversity libraries, fine-tuning the library diversity and making small changes to diversity designs can improve the success rate of single-framework synthetic Fab libraries. These findings have valuable implications for antibody library design, antibody phage display and combinatorial antibody engineering.

5.3. Generation and Validation of Anti-EGFR Fabs from Naïve and Structure-Guided Synthetic Antibody Libraries

In this section, we used EGFR as a model protein to test whether structural information on discontinuous protein interfaces could be used for semi-rationally designing synthetic Fab libraries that are biased towards interacting with a specific site on the receptor. First, we describe the design and construction of a structure-guided Fab library that was biased towards interacting with EGFR domain-II. The structure-guided Fab library was constructed by grafting the EGFR domain-II dimerization arm into CDRH3 and by diversifying key positions within ungrafted CDRs. Next, we describe the isolation, affinity maturation, characterization and epitope-mapping of anti-EGFR Fabs from the structure-guided and naïve synthetic Fab libraries. Screening the structure-guided Fab library against EGFR gave rise to a Fab named DL06, which was affinity-matured to Fab DL06-SR02. For comparison, Library-F was used to generate an anti-EGFR Fab named Fab-H whose binding overlapped with the Fab-DL06. Both Fabs possessed low-nM K_D values for recombinant and cell-surface EGFR and inhibited EGF-mediated EGFR activation. Epitopes of both Fabs mapped to EGFR domains I-II and did not overlap with binding sites of any clinically relevant anti-EGFR antibodies.

5.3.1 Design and Construction of the EGFR Domain II Structure-Guided Fab Library

EGFR contains a 4-domain (I to IV, distal to membrane proximal) modular extracellular ligand-binding domain, a single trans-membrane domain, and intracellular tyrosine kinase and regulatory domains. In the absence of a ligand, the EGFR monomer exists in a dynamic equilibrium between an auto-inhibited conformation, where domains II and IV form an intramolecular interaction, and an extended conformation, where domain II is exposed for dimerization. Upon ligand binding to domains I and III, the equilibrium shifts toward the extended conformation, which promotes receptor dimerization and subsequent activation of downstream signaling pathways (Ogiso *et al*., 2002; Garrett *et al*., 2002; Burgess *et al*., 2003; Ferguson, 2008; Bessman *et al*., 2014) **(Figure 5.19A).** In the ligand-bound extended dimer conformation of EGFR, the homo-dimer interface is composed of multiple discontinuous contacts. All dimer contacts observed in the crystal structures are mediated by domain II. Domain II is divided into 8 disulfide-bonded modules, in which modules 2, 5 and 6 interact with corresponding modules in the neighboring receptor. An antiparallel β-hairpin loop that projects

out from module 5, referred to as the dimerization arm (residues 242-259), makes extensive contacts with the dimerization arm of the neighboring receptor burying a surface area of 855 $Å^2$. In addition to mediating loop-loop interactions, the dimerization arm reaches across the interface to contact domains I and III of the neighboring receptor. Studies have shown that the dimerization arm is necessary but not sufficient EGFR dimerization. Two other smaller interaction sites contribute to the dimer interface, one is at the C-terminus of the dimerization arm (module 6) and the other is close to the N-terminus of domain II (module 2). The total surface area buried in the dimer interface is 1125 $Å^2$ on each receptor (Garrett *et al*., 2002; Burgess *et al*., 2003; Ferguson, 2008). Many reports suggest that the dimer interface extends into domain IV, and domain IV contributes directly to stabilization of the EGFR dimer (Schlessinger, 2002 and Ferguson, 2008).

Based on these observations, we created a structure-guided Fab library by grafting the EGFR domain II dimerization loop into CDRH3 of the Hu4D5-8 Fab framework (Fellouse *et al*., 2007) and by diversifying solvent-accessible residues within CDRL3, CDRH1 and CDRH2 (**Figure 5.19B)**. We hypothesized that the dimerization loop in CDRH3 would direct Fabs in the library towards binding domain II of EGFR, and that other CDRs would optimize this interaction by making additional contributions to binding affinity and specificity. To generate the structure-guided Fab library, we first replaced the CDRH3 apex region of the anti-maltose-binding protein (MBP) Fab phagemid with the 28-residue β-hairpin dimerization loop (C240-C267). The domain II graft was flanked by one NNC codon at each terminus to allow for sequence optimization at graft junctions. CDRH3 anchor residues at the N- and C-termini were not modified. Next, we diversified the solvent-accessible residues within CDRs L3, H1 and H2 of the Hu4D5-8 Fab framework using Kunkel mutagenesis. Diversity designs of these CDRs resembled the design of library-F, a highly validated synthetic Fab library (Persson *et al*., 2013). The domain II structure-guided Fab library contained two fixed CDRs (L1 and L2), three diversified CDRs (L3, H1 and H2) and one CDR containing the EGFR dimerization loop with a theoretical diversity of $\sim 10^{14}$. Following phagemid mutagenesis, the reaction product was transformed into SS320 *E. coli* cells for phage library production, giving a final library diversity of $\sim 5 \times 10^9$. To monitor the quality of the library, we verified graft inclusion and CDR diversification by sequencing 20 phagemids isolated from the library. All phagemids analyzed contained the domain II graft in CDRH3 and varying length and sequence diversity in CDRL3, CDRH1 and CDRH2 (data not shown).

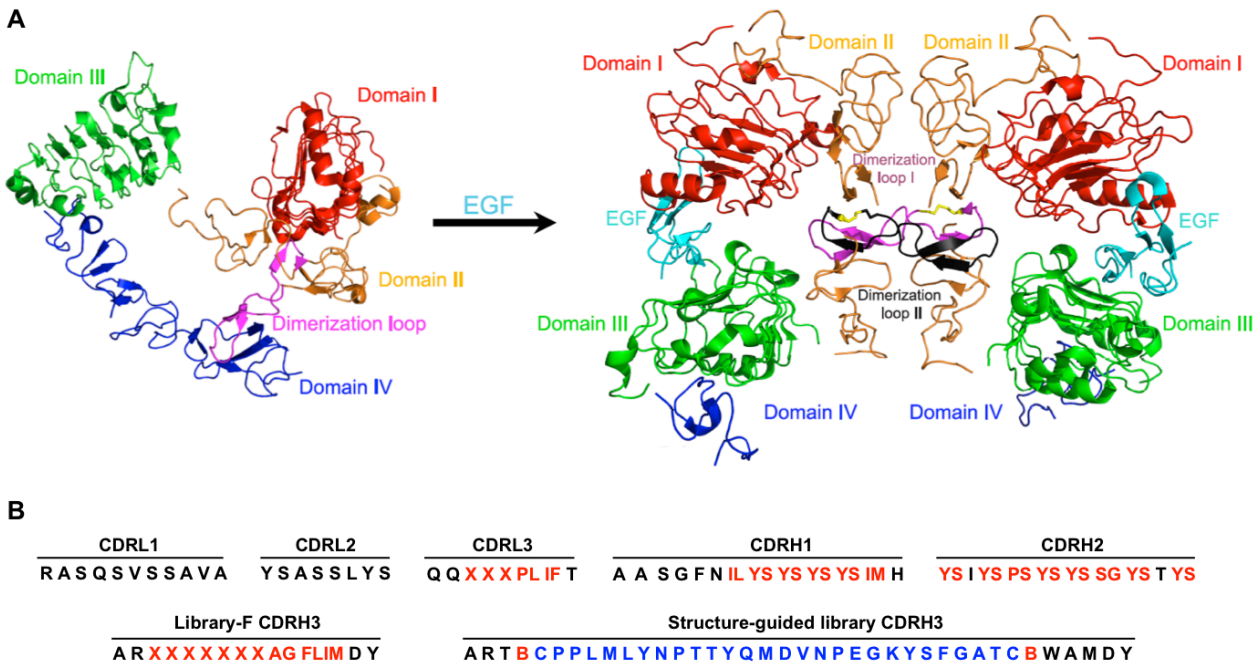


Figure 5.19: Design of EGFR domain II structure-directed Fab library. (A) Structural view of EGF-induced EGFR dimerization. Unliganded EGFR monomer in the tethered conformation (PDB 1NQL) and EGF-bound EGFR dimer in the extended conformation (PDB 1IVO) are shown in cartoon representation. In the absence of a ligand, EGFR predominantly exists in a tethered conformation in which domains II and IV interact and impede ligand binding. Upon EGF binding to domains I and III, EGFR adopts an extended conformation in which domain II is exposed for promoting receptor dimerization. The domain II dimer interface is composed of multiple discontinuous contacts, and a β-hairpin loop in domain II (colored in magenta/black), referred to as the dimerization arm has been shown to be crucial for dimerization. **(B)** CDR diversity designs of library-F and EGFR domain II structure-directed library. Library-F contains two fixed CDRs (L1 and L2) and four diversified CDRs (L3, H1, H2 and H3). Length diversity is included within CDRL3 and CDRH3. X denotes a mixture of nine amino acids (Y, S, G, A, F, W, H, P or V). The design of EGFR domain II structure-guided library is the same as library-F except for CDRH3. CDRH3 of the structure-directed library contains a 28-residue graft corresponding to the EGFR domain II dimerization loop (residues 240-267). B denotes the mixture of amino acids encoded by the NNC codon. The dimerization loop residues in CDRH3 are colored in blue, and the diversified CDR residues in CDRs L3, H1 and H2 are colored in red.

5.3.2 Isolation of EGFR Specific Fabs From Structure-Guided and Naïve Fab libraries

We used the EGFR domain II structure-guided Fab library to conduct *in vitro* solid-phase phage display selections against the recombinant human EGFR ectodomain (EGFR-ECD). Enrichment in target-specific phages was observed after six rounds of selection. Random clone-picking and sequencing analysis of 20 phagemids revealed three unique clones, all containing the dimerization loop graft in CDRH3. Fab-ELISA showed that one clone, named DL06, had the

highest affinity for EGFR-ECD (IC_{50} ~200 nM, data not shown). To improve the affinity of DL06, we constructed a DL06-based affinity maturation library by soft-randomizing the dimerization loop in CDRH3, and used the library for solid-phase selections against EGFR-ECD (see section 4.7). After four rounds, we isolated a Fab named DL06-SR02, which bound to EGFR-ECD with an IC_{50} of ~30 nM (data not shown). Sequencing revealed that DL06-SR02 contained three mutations in the CDRH3 dimerization loop, and one spurious substitution in CDRH1 relative to the parent Fab DL06 **(Figure 5.20A)**. For brevity, DL06-SR02 is referred to as DL06 throughout this report.

To compare whether a naïve Fab library without any grafted motifs is capable of yielding Fabs that recognize domain-II of EGFR, we used library-F to conduct solid-phase selections against EGFR-ECD. Though the domain II structure-guided library and library-F contain the same diversity within CDRs L3, H1, and H2, library-F contains much more diversity within CDRH3 (~10^{17}), whereas the motif-grafted library contains a fixed dimerization loop within CDRH3 flanked by one random residue at each end. The theoretical and actual diversities of library-F were calculated to be ~10^{28} and ~10^{10}, respectively (Persson *et al*., 2013). We isolated several EGFR-specific clones from library-F after three rounds of selection, however we chose to pursue one Fab, named FabH (**Figure 5.20A**), as competitive Fab-ELISA assays revealed that DL06 and FabH competed with each other for binding EGFR-ECD. Competitive Fab-ELISA was performed using Fabs in both non-biotinylated and site-specifically biotinylated formats (BT-Fab). Pre-incubation of EGFR-ECD with saturating concentrations of DL06 inhibited the binding of BT-FabH to EGFR-ECD. As a positive control, binding of the BT-DL06 analog was also inhibited in the presence of DL06 **(Figure 5.20B)**. These results were reproduced in the opposite configuration in which saturating concentrations of FabH was used to pre-block EGFR-ECD, resulting in potent inhibition of both BT-DL06 and the BT-FabH analog positive control **(Figure 5.20C)**.

5.3.3 Characterization DL06 and FabH Interactions With EGFR

To assess the specificity of DL06 and FabH, we performed a Fab-ELISA against immobilized Fc-tagged EGFR and other ErbB family member ectodomains. At a concentration of 50 nM, both Fabs bound specifically to EGFR with no observable binding to other ErbB family members or the Fc protein **(Figure 5.20D)**.

A

Fab	CDRL3	CDRH1	CDRH2	CDRH3
Template	QQSSYSLIT	FSSSSI	SISSSYGYTY	ARTVRGSKKPYFSGWAMDY
DL06	QQWSYYPIT	ISSSSI	YISSYYSSTY	ARTYCPPLMLYNPTTYQMDVNPEGKYSFGATCGWAMDY
DL06-SR02	QQWSYYPIT	**F**SSSSI	YISSYYSSTY	ARTYCPPLM**R**YNPTTYQMDVNPE**AQ**YSFGATCGWAMDY
FabH	QQSYWLIT	LYSSSM	SIYPYSGYTY	ARYPFGVSAYYAMDY

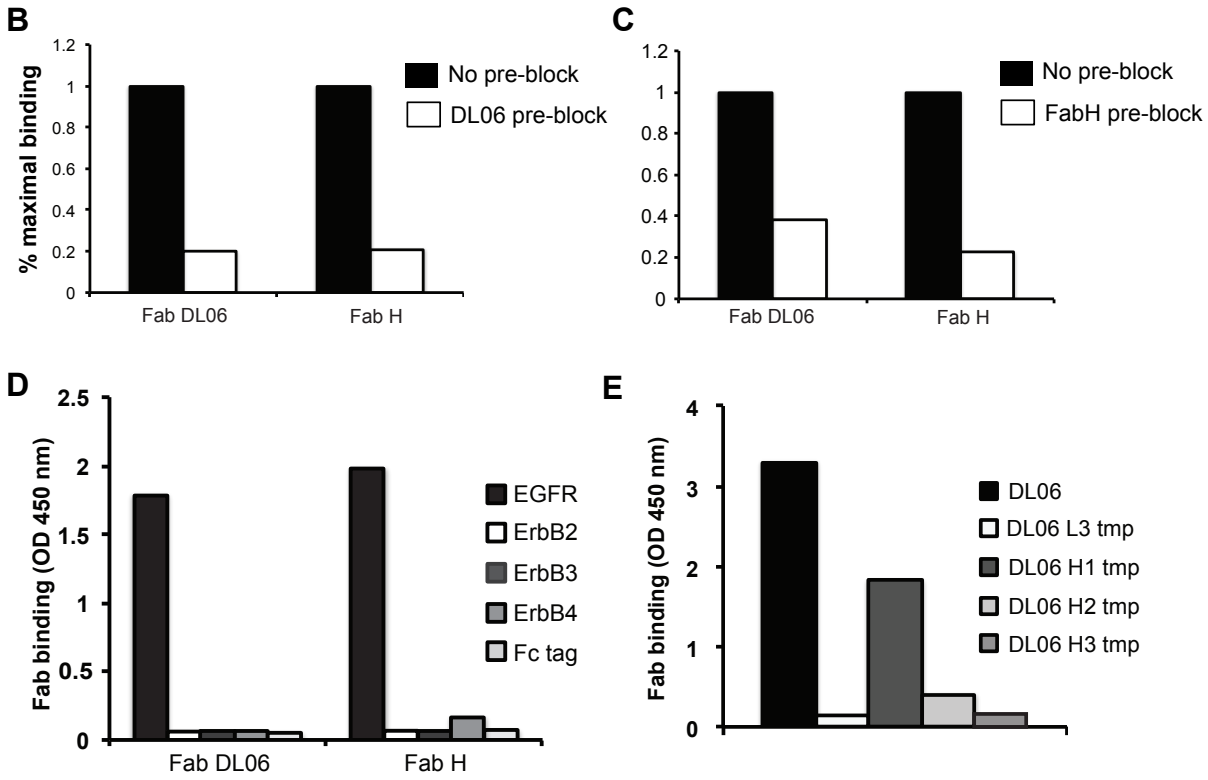
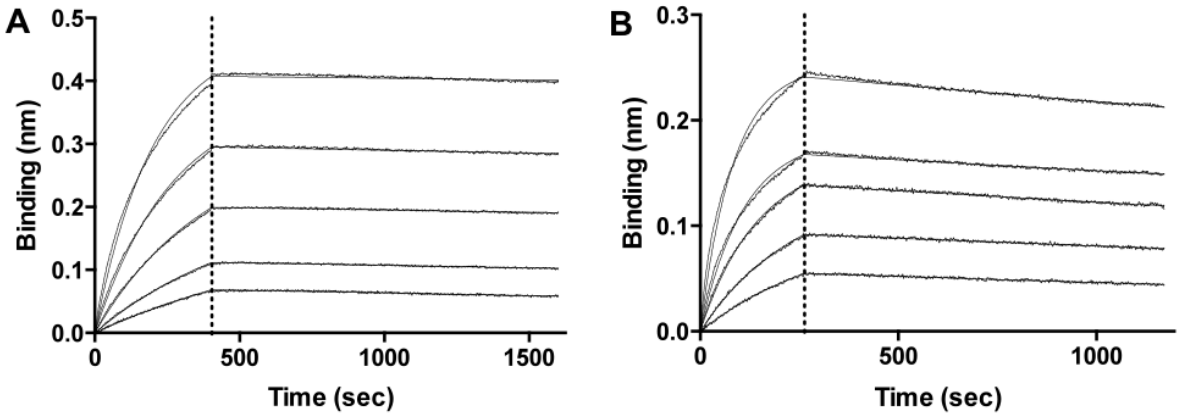


Figure 5.20: Isolation and characterization of Fab-DL06 and Fab-H. (A) CDRL3, CDRH1, CDRH2 and CDRH3 sequences of anti-MBP Fab (template), DL06, DL06-SR02 and Fab-H. Amino acid mutations in DL06-SR02 relative to DL06 are in bold letters. **(B)** Influence of DL06 on BT-DL06 (positive control) and BT-FabH binding to EGFR-ECD assessed by competitive Fab-ELISA **(C)** Influence of FabH on BT-DL06 and BT-FabH (positive control) binding to EGFR-ECD assessed by competitive Fab-ELISA. In panels B and C, Fab-ELISA was performed using HRP-conjugated streptavidin. **(D)** Specificity of Fab-DL06 and Fab-H towards ErbB family members assessed by single-point Fab-ELISA at 50 nM Fab concentration. **(E)** Effect of CDR perturbations in DL06. Diversified CDR sequences in DL06 were back mutated to anti-MBP template Fab sequences, and purified DL06 CDR mutants were tested for binding to EGFR-ECD by single-point Fab-ELISA at 200 nM Fab concentration. In panels D and E, Fab-ELISA was performed using an HRP-conjugated anti-FLAG secondary antibody.



C

Fab	Target	K_{ON} (1/Ms)	K_{ON} error (1/Ms)	K_{OFF} (1/s)	K_{OFF} error (1/s)	K_D (M)	K_D error (M)
DL06	EGFR	7.2E+03	41	5.6E-05	4.4E-07	6.8E-09	1.0E-10
FabH	EGFR	1.3E+05	567	1.7E-04	7.3E-07	1.4E-09	1.1E-11

Figure 5.21: Binding kinetic analysis of anti-EGFR Fabs. Bio-layer interferometry was used for analyzing the binding of EGFR-ECD to sensor-immobilized DL06 **(A)** and FabH **(B).** Association and dissociation were monitored using serial two-fold dilutions of EGFR-ECD, starting from 1480 nM EGFR-ECD for DL06 and 200 nM EGFR-ECD for FabH. Data points were globally fit to 1:1 binding model. **(C)** Binding parameters obtained for Fab-EGFR interactions.

To confirm that the CDRH3 EGFR domain II graft in DL06 contributes to Fab binding and to measure relative contributions of other diversified CDRs towards Fab-EGFR interactions, we back-mutated sequences of each diversified CDR in DL06 to the corresponding anti-MBP Fab CDR and evaluated the binding of these Fabs to EGFR. A Fab-ELISA assay conducted on EGFR-ECD using 200 nM DL06 or the various CDR variants revealed that binding was substantially diminished by changing CDRs L3, H2, or H3 to anti-MBP Fab CDRs, whereas reversion of CDRH1 to anti-MBP Fab CDRH1 resulted in only a ~ 45% loss of binding **(Figure 5.20E)**. This demonstrated that the CDRH3 graft was required for Fab binding, and also indicated that other diversified CDRs made critical contributions to Fab-EGFR interactions.

To measure kinetics of DL06 and FabH binding to EGFR-ECD, we used bio-layer interferometry. Fabs were immobilized on amine-reactive sensors and exposed to increasing concentrations of EGFR-ECD. Association and dissociation rates were assessed by a wavelength shift and kinetic data sets were globally fit using a 1:1 binding model **(Figures 5.21A and 5.21B)**. The kinetic parameters for Fab-EGFR interactions are summarized in **Figure 5.21C**.

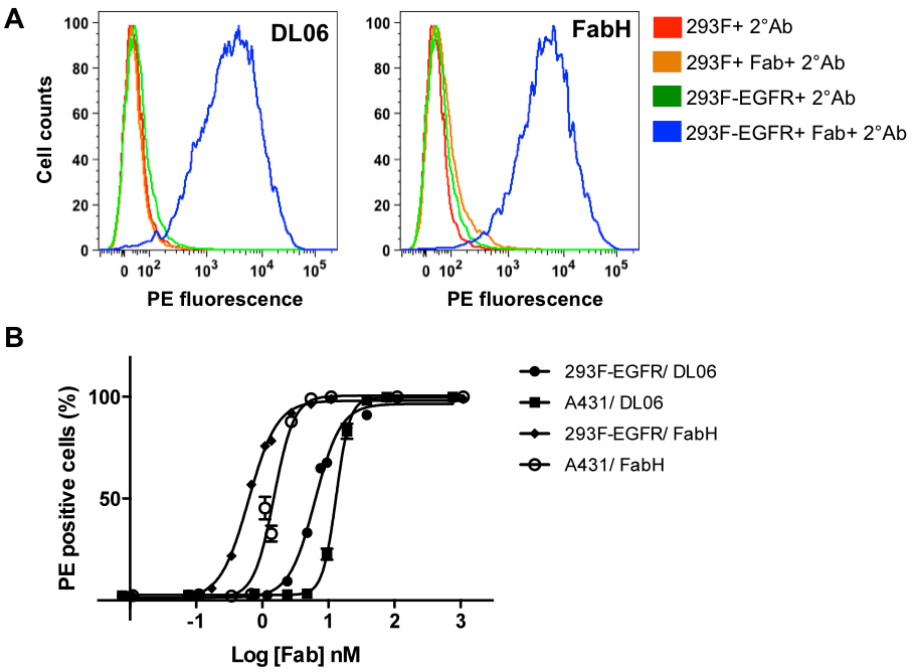


Figure 5.22: Analysis of anti-EGFR Fabs binding to cell-surface EGFR. (A) Binding of DL06 and FabH to untransfected 293F cells and 293F cells-expressing WT EGFR was assessed by flow cytometry using a PE-conjugated anti-FLAG secondary antibody. **(B)** Determination of EC_{50} values for DL06 and FabH binding to 293F cells-expressing WT EGFR and A431 cells endogenously expressing EGFR. Binding was assessed by flow cytometry using a PE-conjugated anti-FLAG secondary antibody. The frequency of PE-positive cells was determined for a range of Fab concentrations, and data points were fit to one-site specific binding model.

DL06 and FabH bound to EGFR-ECD with K_D values of 6.8 ± 0.1 nM and 1.4 ± 0.01 nM, respectively. DL06 bound ~5-fold weaker than FabH, with the main difference being that DL06 had a slower K_{ON}.

To check whether DL06 and FabH bind to cell-surface EGFR, we used flow cytometry. We overexpressed a full-length EGFR construct fused to GFP at its C-terminus (EGFR-GFP) (Carter and Sorkin, 1998) in HEK293F cells, and assessed flag-tagged Fab binding to HEK293F and EGFR-expressing HEK293F cells using a PE-conjugated mouse anti-FLAG antibody. At a concentration of 100 nM, >95% of EGFR-expressing cells were PE-positive, indicating that DL06 and FabH bound to cell-surface EGFR, and no significant binding was observed with untransfected HEK293F cells **(Figure 5.22A)**. To measure EC_{50} values for Fabs binding to cell-surface EGFR, we incubated various concentrations of DL06 and FabH with EGFR-expressing HEK293F and A431 epidermoid carcinoma cells, and plotted the frequency of PE-positive cells

versus Fab concentration **(Figure 5.22B)**. Both Fabs exhibited saturation behavior and low-nM EC_{50} values for cell-surface EGFR. DL06 bound to 293F-EGFR and A431 cells with EC_{50} values of 8.15 ± 0.4 nM and 12.12 ± 1.5 nM, respectively. FabH bound to 293F-EGFR and A431 cells with EC_{50} values of 0.57 ± 0.02 nM and 1.49 ± 0.2 nM, respectively. FabH bound 14-fold stronger to 293F-EGFR and 8-fold stronger to A431 cells than DL06.

5.3.4 Mapping DL06 and FabH Interaction Domains on EGFR

To confirm that DL06 and FabH did not bind to previously characterized epitopes in EGFR domain III, we tested the influence of two monoclonal antibodies (mAbs), cetuximab (CTX) and panitumumab (PMB), on DL06 and FabH binding to EGFR. CTX and PMB bind exclusively to EGFR domain III. Their epitopes substantially overlap with each other, and partially overlap with the EGF-binding site in EGFR domain III (Voigt *et al*., 2012). Pre-incubation of EGFR-ECD or cell-surface EGFR with saturating concentrations of CTX did not affect the binding of DL06 and FabH to EGFR **(Figures 5.23A and 5.23B)**, indicating that epitopes of CTX and DL06 and FabH are distinct from each other. Pre-incubation of EGFR-ECD or cell-surface EGFR with saturating concentrations of PMB partially blocked DL06 and FabH binding to EGFR **(Figures 5.23C and 5.23D)**, suggesting that epitopes of PMB and DL06 and FabH could partially overlap with each other. To confirm this, we used phage-ELISA to test whether phage-displayed EGFR domain III (Tundidor *et al*., 2014) could bind to immobilized DL06 and FabH. No binding was observed between Fabs and EGFR domain III, indicating that domain III alone was not sufficient for EGFR-Fab interactions. As a positive control, EGFR domain III bound to CTX and PMB in a dose-dependent manner **(Figure 5.23E)**. The partial blockage of DL06 and FabH binding by PMB may be caused by steric interactions between Fabs and PMB binding sites, and/or an allosteric influence of PMB on Fab-binding sites.

To conduct domain-level epitope mapping of DL06 and FabH, we generated EGFR-GFP truncation constructs encoding individual domains or their combinations, and tested the binding of DL06, FabH and control mAbs (CTX and PMB) to EGFR-GFP truncations expressed on HEK293F cells. GFP-fluorescence was used to monitor EGFR expression and PE-conjugated secondary antibody was used to monitor Fabs or mAbs binding to cells. Binding of Fabs and mAbs to EGFR truncations is summarized as a comparative heat-map in **Figure 5.24**. We expected that CTX and PMB could be used as positive controls for all domain-III containing constructs. However, both CTX and PMB did not bind to D3, D2-D3, and D1-D2-D3, indicating

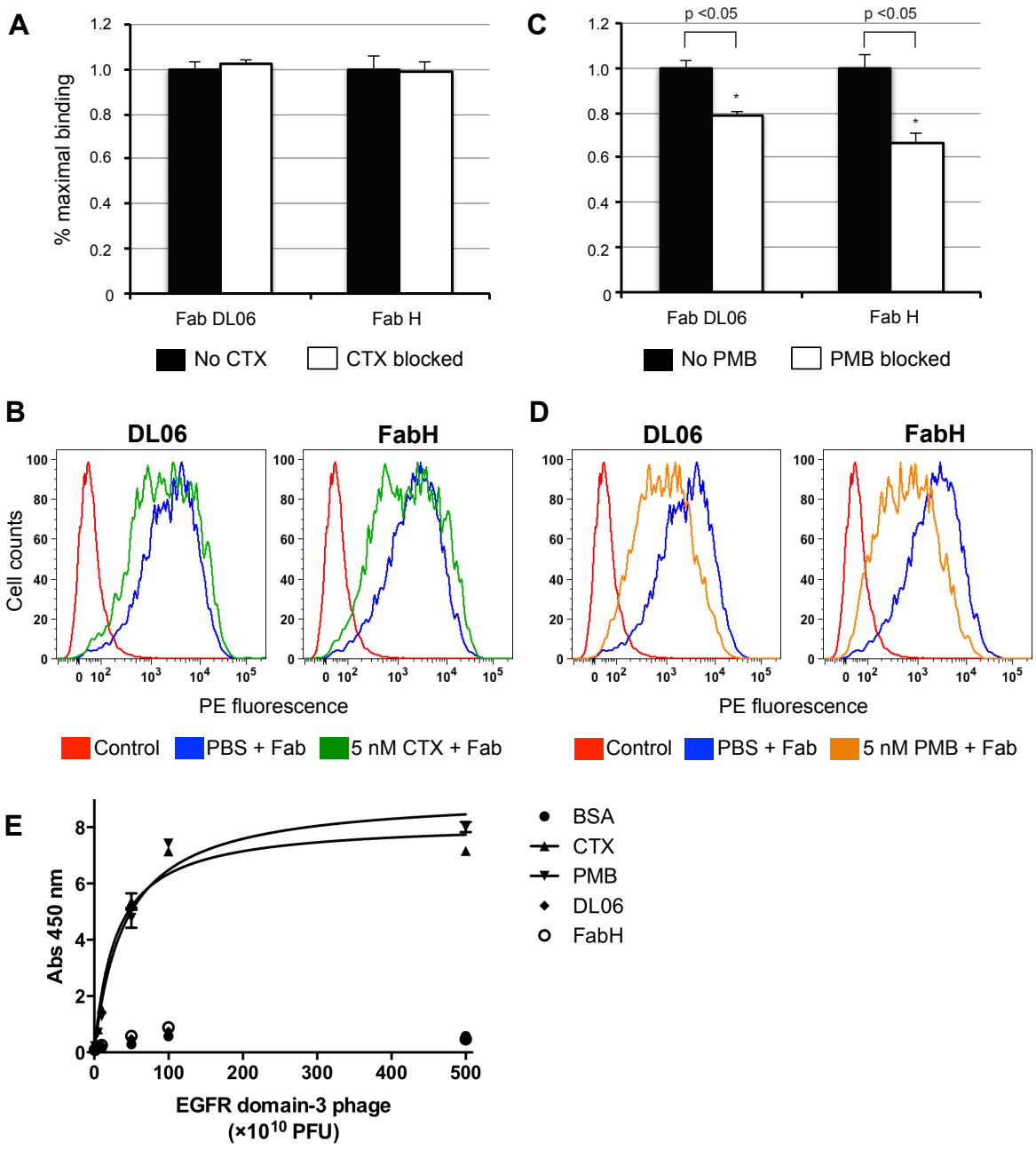


Figure 5.23: Influence of anti-EGFR mAbs on anti-EGFR Fabs binding to EGFR. (A) Influence of CTX on DL06 and FabH binding to EGFR-ECD. **(B)** Influence of CTX on DL06 and FabH binding to cell-surface EGFR. **(C)** Influence of PMB on DL06 and FabH binding to EGFR-ECD. **(D)** Influence of PMB on DL06 and FabH binding to cell-surface EGFR. Influence of mAbs (CTX and PMB) on binding of Fabs (DL06 and FabH) to EGFR-ECD was assessed by competitive Fab-ELISA using an HRP-conjugated anti-FLAG secondary antibody. Influence of mAbs (CTX and PMB) on binding of Fabs (DL06 and FabH) to cell-surface EGFR was assessed by flow cytometry using a PE-conjugated anti-FLAG secondary antibody. **(E)** Analysis of phage-displayed EGFR domain III binding to immobilized BSA, CTX, PMB, DL06 and FabH by phage-ELISA using an HRP-conjugated anti M13 (phage coat protein) secondary antibody.

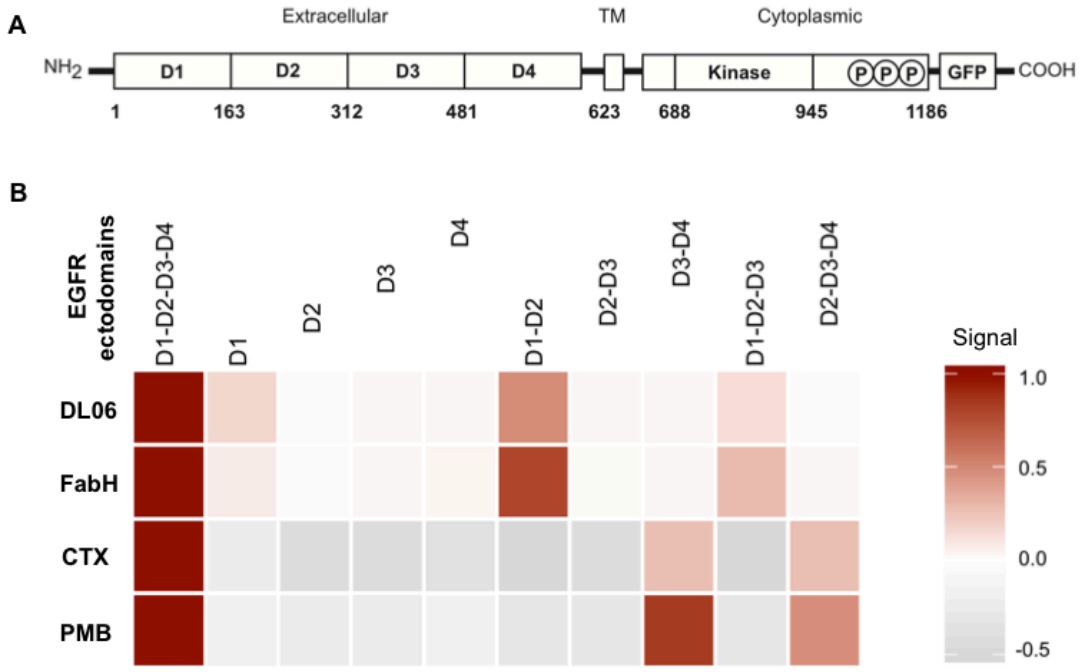


Figure 5.24: Domain-level epitope mapping of anti-EGFR Fabs. (A) Schematic of domain organization for the WT EGFR-GFP construct. **(B)** Heatmap showing the binding of anti-EGFR Fabs (DL06 and FabH) and mAbs (CTX and PMB) to WT EGFR and EGFR truncations expressed on 293F cells. Binding was assessed by flow cytometry using a PE-conjugated anti-FLAG secondary antibody for DL06 and FabH, and a PE-conjugated anti-human IgG secondary antibody for CTX and PMB. A heatmap was constructed from PE-positive cell signals to visualize and compare relative Fab/mAb binding to EGFR and mutant-transfected 293F cells by subtracting signals from untransfected cells and normalizing to maximal WT-EGFR signals.

that some truncations had unstable domain breakpoints for proper folding or expression. Both CTX and PMB bound to D3-D4 and D2-D3-D4, indicating that N-terminus truncations were tolerated by EGFR. DL06 and FabH interacted with three truncation constructs (D1, D1-D2, and D1-D2-D3), however the binding was less than that observed with WT-EGFR. DL06 and FabH interacted most strongly with the D1-D2 construct. Deletion of D1 or D1-D2 from EGFR (constructs D2-D3-D4 and D3-D4) completely abolished DL06 and FabH binding. In contrast, CTX and PMB interacted with D2-D3-D4 and D3-D4. Together, this indicated that epitopes for DL06 and FabH were localized within domains I and II. DL06 bound ~2-fold tighter to D1 than FabH, whereas FabH bound ~2-fold tighter to D1-D2 than DL06, suggesting that both Fabs targeted different epitopes within domains I and II. While CTX and PMB did not bind to D1-D2-D3, Fabs exhibited weaker binding to this construct relative to D1-D2. This could have resulted

from the changes in overall structural orientations or conformations in domains I to III upon deleting domain IV. Fabs and mAbs did not bind to D2 and D4 constructs, but no conclusions were drawn because of the absence of a positive control for binding.

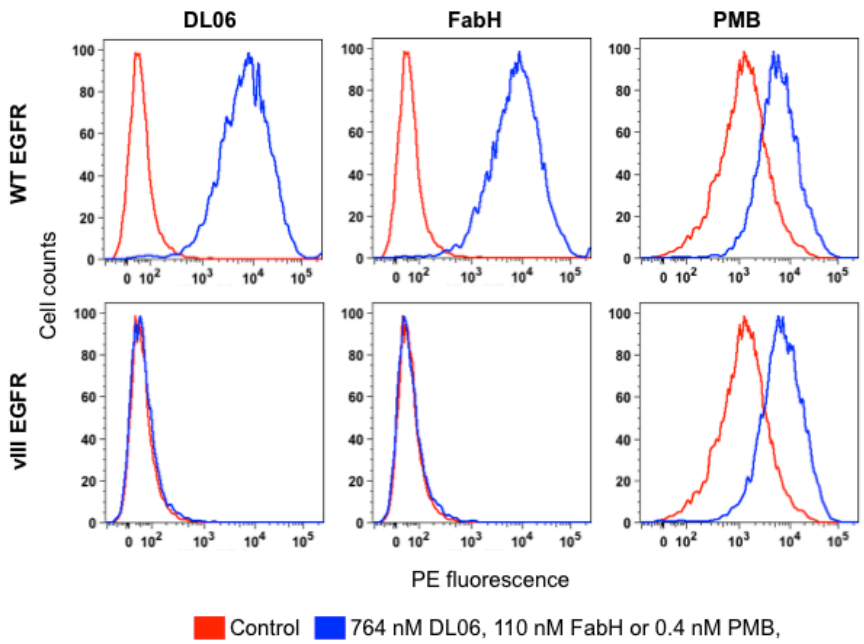
To confirm the importance of EGFR domain II in Fab-EGFR interactions, we assayed DL06, FabH, and PMB binding to the EGFR-vIII glioblastoma mutant expressed on HEK293 cells using flow cytometry. In EGFR-vIII, EGFR amino acids 6-274 are replaced by a glycine, which leads to the deletion of domain-I and 75% of domain-II, including the dimerization loop (Gan *et al*., 2009). As expected, both DL06 and FabH did not bind to EGFR-vIII, whereas PMB bound to EGFR-vIII **(Figure 5.25A)**. To test the importance of EGFR dimerization loop on DL06 and FabH binding, we assayed DL06, FabH, and PMB binding to two dimerization loop mutants (D28-EGFR and DM-EGFR) expressed on HEK293 cells using flow cytometry. D28-EGFR contains a deletion of 28 amino acids corresponding to the dimerization loop (ΔC240-C267), and DM-EGFR has two point mutations (Y251A and R285S) in domain II. Both constructs have been shown to prevent EGFR homo-dimerization (Ogiso *et al*., 2002 and Dawson *et al*., 2005). The EC_{50} values for DL06 and FabH binding to D28-EGFR and DM-EGFR were only ~3-fold weaker relative to WT-EGFR. PMB showed a ~2-fold decrease in binding to D28-EGFR, and exhibited similar EC_{50} values for binding to WT-EGFR and DM-EGFR **(Figure 5.25B)**. These results were surprising for DL06, since we believed that the dimerization loop graft in CDRH3 would target binding to the dimerization loop of EGFR, and thus, deletion of the dimerization loop in EGFR was expected to have a much greater effect on Fab binding.

5.3.5 DL06 and FabH Inhibit EGF-Mediated EGFR Activation

To check the influence of EGF on DL06 and FabH binding to cell-surface EGFR, we incubated EGFR-expressing HEK293F cells with EGF for 15 minutes, and measured EC_{50} values for DL06 and FabH binding to EGF-treated cells using flow cytometry **(Figures 5.26A and 5.26B)**. At 10 nM EGF concentration, DL06 and FabH had EC_{50} values of 127.05 ± 12.62 nM and 2.86 ± 0.15 nM, respectively. In comparison with EC_{50} values for EGF-untreated cells, DL06 showed a ~16-fold reduction in binding, and FabH showed a ~5-fold reduction in binding. At 30 nM EGF concentration, DL06 titrations did not show saturation therefore EC_{50} was not calculated. With an EC_{50} of 11.25 ± 0.12 nM, FabH showed ~18-fold reduction in binding

relative to EGF-untreated cells. The results indicated that EGF had a stronger effect on DL06 than FabH binding to cell-surface EGFR. Given that DL06 showed partial binding to EGFR domain I, the strong EGF effect suggested that DL06 could partially overlap with the EGF-binding site in domain I. EGF-induced reduction in Fab binding could also have resulted from the change in EGFR conformation.

A



B

	DL06	FabH	PMB
	EC_{50} (nM)	EC_{50} (nM)	EC_{50} (pM)
WT EGFR	8.15 ± 0.4	0.57 ± 0.02	43.03 ± 1.12
D28 EGFR	21.7 ± 2.7	1.45 ± 0.05	74.54 ± 2.80
DM EGFR	22.85 ± 2.1	1.37 ± 0.07	44.29 ± 0.63

Figure 5.25: Effect of EGFR domain II mutations on binding of anti-EGFR Fabs to cell-surface EGFR. (A) Analysis of DL06, FabH and PMB binding to WT EGFR and vIII EGFR by flow cytometry. Binding of DL06 and FabH to 293F cells and WT- or vIII-expressing 293F cells was assessed using a PE-conjugated anti-FLAG secondary antibody. Binding of PMB to 293F cells and WT- or vIII-expressing 293F cells was assessed using a PE-conjugated anti-human IgG secondary antibody. **(B)** EC_{50} values for DL06, FabH and PMB binding to WT EGFR, D28 EGFR and DM EGFR. EGFR constructs were expressed on HEK293F cells, and binding was assessed by flow cytometry using PE-conjugated antibodies mentioned in panel A. The frequency of PE-positive cells was determined for a range of Fab and PMB concentrations, and data points were fit to one-site specific binding model. EC_{50} values are summarized in panel B.

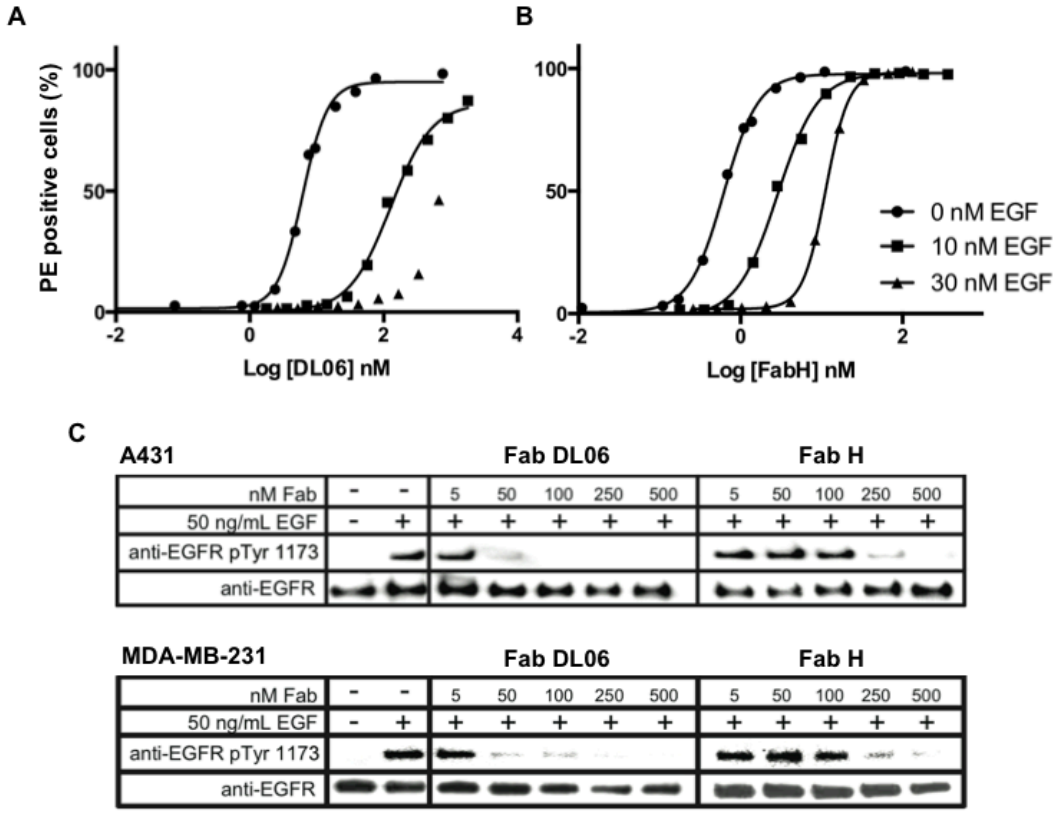


Figure 5.26: Fabs inhibit EGF-mediated EGFR activation. (A) Influence of EGF on DL06 binding to EGFR expressed on HEK293F cells. **(B)** Influence of EGF on FabH binding to EGFR expressed on HEK293F cells. Binding of Fabs to cell-surface EGFR in the presence and absence of EGF was assessed by flow cytometry using a PE-conjugated anti-FLAG secondary antibody. The frequency of PE-positive cells was determined for a range of Fab concentrations, and data points were fit to one-site specific binding model. **(C)** Western blots showing the inhibition of EGF-induced EGFR phosphorylation by DL06 and FabH in A431 and MDA-MB-231 cells. Western analysis was conducted using rabbit polyclonal EGFR and EGFR pTyr1173 primary antibodies and HRP-conjugated anti-rabbit secondary antibodies.

To evaluate consequences of DL06 and FabH binding on EGF-induced EGFR activation, we used two EGFR over-expressing cell lines: A431 epidermoid carcinoma cells and MDA-MB-231 breast cancer cells. Serum-starved cells were incubated with various Fab concentrations for 60 minutes, and then stimulated with 50ng/mL EGF for 15 minutes. Western analysis was performed to detect EGFR phosphorylation at Tyr1173 and total EGFR. Western blots showed the dose-dependent inhibition of EGF-induced EGFR activation by DL06 and FabH **(Figure 5.26C)**. Saturating levels of DL06 and FabH inhibited >95% of EGFR phosphorylation without altering the total EGFR content. Even though DL06 exhibited lower affinities for recombinant

EGFR and cell-surface EGFR, and experienced a higher EGF-induced reduction in binding to cell-surface EGFR, DL06 showed ~5-fold higher inhibition of EGFR activation than FabH, suggesting that small differences in Fab epitopes could lead to Fabs with different mechanisms of action.

5.3.6 Discussion

Structures of protein complexes provide a wealth of information for designing antibodies that mimic and disrupt protein interactions. The combination of synthetic antibody libraries and combinatorial phage display enables precise control of sequences that mediate interactions between antibodies and target proteins and provides an unprecedented opportunity for engineering antibody properties by directed evolution. The purpose of this work was to test whether structural information on discontinuous protein interfaces could be used for semi-rationally designing synthetic Fab libraries that are biased towards interacting with a desired site on the target receptor. Our work builds on previous studies showing the ability of antibodies to accept discrete motifs and small domains as grafts within their CDRs.

We used structural information on the EGFR homo-dimerization interaction (Garrett *et al*., 2002; Burgess *et al*., 2003) to design a structure-guided Fab library that was biased towards interacting with domain II of EGFR. To construct this Fab library, we grafted the EGFR domain II dimerization arm into CDRH3. This graft alone was not capable of interacting with EGFR-ECD. Since EGFR homo-dimerization involves multiple discontinuous regions that cover a large surface area (1125 $Å^2$), it is not practical to graft an entire EGFR homo-dimerization interaction into a CDR. Antibody interactions however can cover a surface area of 2071 ± 456 $Å^2$ using several CDRs (Ramaraj *et al*., 2012), and thus we diversified additional CDRs (L3, H1 and H2) in the Fab framework. We hypothesized that by grafting a receptor interaction motif into CDRH3, combined with diversification of additional CDRs, we could isolate Fabs with specificity for the native interaction domain in the target receptor. By targeting EGFR domain II by design, the aim was to obtain lead antibodies that would block receptor activation and dimerization, but whose mechanism of action would be distinct from the existing therapeutic antibodies. We further sought to compare the grafting strategy with selections using a functional naïve synthetic Fab library.

Solid-phase phage display screens led to the isolation of EGFR-specific Fabs DL06 and FabH from structure-guided and naïve synthetic Fab libraries, respectively. Six selection rounds

were required to isolate DL06 from the structure-guided library. Further, the affinity of DL06 was lower than FabH isolated from the naïve synthetic Fab library. These results suggested that high-affinity binding clones are relatively rare in the structure-guided library. Affinity maturation was required to improve the affinity of DL06, whereas FabH isolated from Library-F was of suitably high affinity. Both Fabs possessed low-nM affinities for EGFR-ECD and cell-surface EGFR. DL06 isolated from the structure-guided library required the domain II dimerization loop graft in CDRH3 for EGFR binding, but additionally required each of the remaining CDRs (L3, H2, and to a lesser degree- H1).

Domain-level epitope mapping indicated that epitopes for DL06 and FabH span across multiple extracellular domains of EGFR. Among the EGFR truncation constructs, DL06 and FabH interacted most strongly with the D1-D2 construct. In addition to making a direct interaction with Fabs, domain I may also be involved in proper positioning of domain II for Fab binding. Glioblastoma mutations in the domain I-domain II interface have been shown to directly influence the domain II conformation (Bessman *et al*., 2014). Domains III and IV may contain partial Fab epitopes or may be involved in proper orientation of the EGFR conformation required for Fab binding. DL06 and FabH competed with each other for binding to EGFR-ECD, confirming substantial overlap of their epitopes during domain mapping. Fab/EGF competition studies indicated that Fab epitopes are conformation dependent and are lost upon EGF binding to EGFR. Despite the fact that Fabs interacted with D1-D2, deletion of the EGFR domain II dimerization loop did not show significant changes in Fab binding. Relative to WT EGFR, Fabs only showed a 3-fold decrease in binding to D28 EGFR. This limited loss in binding indicates that there are difficulties in precisely controlling the targeting epitopes of Fabs when using grafting strategies for mimicking discontinuous protein-protein interactions. The ungrafted CDRs may play a more dominant role in directing binding interactions than the grafted CDR, and perhaps a limited diversification strategy could be effective in ungrafted CDRs for obtaining Fab clones that bind to a specific motif on a receptor.

The binding of an inhibitor to the domain II dimerization arm of EGFR has not yet been described, however such an agent would clearly be a potent inhibitor of ligand-induced dimerization of EGFR. It is possible that the lack of such an inhibitor may result from the relatively limited selection procedures that have been used to generate the current repertoire of inhibitory antibodies. Anti-EGFR antibodies in the clinic target domain III of EGFR, which is

likely due to its preferred tethered conformation used to generate antibodies. This is in contrast to ErbB2, which exists predominantly in the extended conformation, where antibodies in the clinic target domains II (pertuzumab) and IV (trastuzumab) (Fauvel and Yasri, 2014). Immunizing mice with NR6 fibroblasts expressing the EGFR vIII mutant (Δ6-273) generated EGFR domain II antibodies (Mab806 and Mab7A7) that are currently under clinical development (Gan *et al*., 2012). These antibodies bind to a linear epitope in domain II (residues 287-302), which is located at the C-terminus of the dimerization loop. Mab806 binds to EGFR-vIII and overexpressed WT-EGFR in the extended monomer conformation (Garrett *et al*., 2009). Studies have shown that domain III and domain II EGFR antibodies possess different mechanisms of action. At the receptor level, domain III antibodies prevent ligand binding to EGFR and sterically inhibit the formation of the extended monomer conformation whereas domain II antibodies prevent EGFR homo- or hetero-dimerization (Schmitz and Ferguson, 2009; Gan *et al*., 2012). Phase-I clinical trials with Mab806 showed excellent targeting of tumor sites without any significant toxicity (Scott *et al*., 2007), thus it will be interesting to explore the activity of DL06 and FabH antibodies against EGFR-overexpressing cancer cells and mouse models.

Our studies clearly distinguished the epitopes of DL06 and FabH from all other antibodies targeting EGFR. Cetuximab, panitumumab and nimotuzumab bind exclusively to domain III (Schmitz and Ferguson, 2009), whereas domain III alone was not sufficient for DL06 and FabH interactions. Further, the presence of cetuximab did not affect the binding of DL06 and FabH to EGFR. Mab806 binds to a linear epitope in domain II and binds stronger to EGFR-vIII than to WT-EGFR (Garrett *et al*., 2009), whereas DL06 and FabH bind to a complex epitope spanning multiple EGFR domains and do not bind to EGFR-vIII. The continued development of anti-EGFR antibodies has been driven by the observation that targeting novel epitopes can provide potential clinical benefits. First, targeting an epitope distinct from that of a primary therapy can circumvent resistance mutations that render first line therapies ineffective (Montagut *et al*., 2012). Second, combinations of non-competitive antibodies can offer additional potency outside of immune-effector functions of antibodies by enhancing receptor inactivation (Klapper *et al*., 1997) or by promoting receptor internalization (Friedman *et al*., 2005) and are proving clinically useful (Yamashita-Kashima *et al*., 2011; Ko *et al*., 2015). DL06 and FabH could prove beneficial using both the mechanisms, as they target unique epitopes that are different from previously validated epitopes on EGFR.

Despite the apparent overlap of DL06 and FabH epitopes, we observed clear differences in their *in vitro* binding and cell-based activity. During binding kinetic analysis, DL06 displayed an 18-fold slower K_{ON} and a 3-fold slower K_{OFF} than FabH, suggesting that DL06 and FabH could use different binding mechanisms to interact with EGFR. Even though DL06 exhibited lower affinities for recombinant EGFR and cell-surface EGFR, DL06 showed ~5-fold higher inhibition of EGFR phosphorylation than FabH. Fab/EGF competition studies indicated that EGF had a stronger effect on DL06 than FabH binding to cell-surface EGFR. These results suggest that DL06 and FabH could use different mechanisms of action to prevent EGFR activation. Although it was possible to obtain similar Fabs from both libraries, in light of noted differences it is difficult to conclude that one approach offers a stark advantage over the other. Rather, we conclude that both approaches may offer alternatives to conventional antibody discovery techniques.

In summary, we generated and validated two anti-EGFR Fabs from structure-guided and naïve synthetic Fab libraries. Both Fabs possessed low-nM K_D values for recombinant and cell-surface EGFR, recognized new epitopes on EGFR, and inhibited EGF-mediated EGFR activation. Though we are not the first to obtain anti-EGFR antibodies from wholly synthetic sources (Schaefer *et al*., 2011), our work reconfirms the potential of single-framework synthetic Fab libraries in generating novel antibody leads for clinical development. Synthetic libraries, whether based upon naïve or structure-directed diversity, proved advantageous over conventional immunization techniques by allowing access to valid untargeted epitopes on a clinically-relevant drug target. Further, our work highlights the utility of structure-guided synthetic libraries and the versatility of naïve synthetic libraries in obtaining specific, potent and novel antibodies.

6. CONCLUSIONS AND FUTURE DIRECTIONS

Synthetic antibody technology has been successful in generating antibodies against a wide variety of antigens with desirable biophysical, biochemical, cell biological, pharmacological and clinical properties. As of July 2016, 13 synthetic antibody-based drugs (12 IgGs and one IgG-drug conjugate) have been undergoing phase II clinical studies and three synthetic antibodies (IgGs) have been undergoing phase III clinical studies. These 16 synthetic antibodies were derived from phage-displayed synthetic Fab or scFv libraries (Frenzel *et al*., 2016).

Due to the limitations in the quality of the phage-displayed antibody library (Ponsel *et al*., 2011; Zhai *et al*., 2011), the immobilized target (Koide *et al*., 2009; Miersch and Sidhu, 2012) and the phagemid/phage/host system (Derda *et al*., 2011; Saggy *et al*., 2012), phage display screening of antibody libraries against immobilized targets does not give rise to antibody fragments in all selection experiments. Further, antibody fragments discovered from a successful selection may not possess favorable production, biological and pharmacological characteristics for drug development (Beck *et al*., 2010; Ponsel *et al*., 2011; Nixon *et al*., 2014). Therefore, special strategies are required for increasing the success rate of phage display selections and for obtaining useful antibodies with desirable properties. In this thesis, we have implemented various strategies for generating high-quality Fabs against Notch, Jagged and EGF receptors.

We designed, constructed and validated a new phage-displayed single-framework synthetic Fab library, named library-S. The library was built on a modified Hu4D5 framework that contained three point mutations in the FRM3 region to accommodate the new CDRH2 loop. Four CDRs were fixed to preserve the most-frequent canonical CDR conformation preferred by the chosen Fab framework. Length, amino acid, and conformational diversity were engineered within CDRL3 and CDRH3 using custom-designed tri-nucleotide mixes. During CDR diversification, template retention was substantially reduced by restriction enzyme digestion of the phagemid library. The quality of the phage library was assessed in terms of overall sequence composition, length, and amino acid distributions. The library contained ~8.5 billion unique Fabs and >95% of the library correctly encoded both the diversified CDR sequences. The library was shown to be functional by conducting solid-phase selections against Notch and Jagged receptors. Further studies are required to characterize the effects of new CDRs and framework substitutions

on the biophysical properties of the modified Hu4D5-8 framework, and to fully explore the potential of the library against different classes of targets.

In recent years, NGS technologies have been employed in several phage display platforms for analyzing natural and synthetic antibody fragment sequences, and for identifying and reconstructing scFv and Fab binders not found by conventional ELISA screens (Naso *et al*., 2014; Glanville *et al*., 2015). In this work, we developed an NGS-assisted antibody discovery platform by integrating phage-displayed single-framework synthetic Fab libraries with Ion Torrent sequencing. Due to limitations in attainable read lengths, NGS analysis of Fab libraries and selection outputs is usually restricted to CDRH3. Since CDRH3 information alone is not sufficient for high-throughput reconstruction of Fabs, we developed a method for linking and sequencing all diversified CDRs in phage Fab pools without losing right CDR combinations. Our method determined the complete fingerprint of selection outputs and paved the way for reliable and straightforward conversion of NGS information into Fab clones. The method is rapid and simple and can be adapted to any synthetic scaffold-based phage library. We used our NGS-assisted Fab reconstruction method to recover low-frequency rare clones from phage selection outputs. While previous studies chose rare clones for rescue based on their relative frequencies in sequencing outputs (Ravn *et al*., 2010; D'Angelo *et al*., 2014), we chose rare clones for reconstruction from less-frequent CDRH3 lengths. In some cases, reconstructed rare clones (frequency ~0.1%) showed higher affinity and better specificity than high-frequency top clones isolated by Sanger sequencing, highlighting the significance of NGS-based approaches in synthetic antibody discovery.

In this work, NGS information was useful for (1) assessing the quality of Fab libraries, (2) analyzing selection outputs in terms of sequence evolution, length and amino acid distribution, and (3) identifying and reconstructing low-frequency rare clones from selection outputs. Most notably, we used NGS information for identifying and reconstructing rare clones with less-frequent CDRH3 lengths, for studying the pairing between CDRL3 and CDRH3 lengths in the selection output, and for fine-tuning the library length diversity. These advances in the use of NGS information gave rise to Fab clones with better binding properties. We used CDRH3 length as a parameter for designing and screening libraries and for isolating Fabs from selection outputs. This approach proved to be successful for obtaining Fabs with different and superior binding profiles.

In this work, we reconstructed rare Fab clones in the phagemid vector and sub-cloned the ELISA-positive clones into the Fab expression vector. In future, desired CDR combinations can be reconstructed directly in suitable Fab/scFv/IgG expression vectors, and tested for binding to relevant antigens using bio-layer interferometry in a high-throughput format. The NGS approach described in this work can be used for identifying and reconstructing Fabs from biopanning experiments where isolating target-specific Fabs is more difficult than from solid-phase selections (Tomic *et al*., 2015). NGS analysis can be used for comparing CDR sequences from negative and positive selections, and enriched CDR combinations in target phage pools can be used for reconstruction. In this work, we obtained only ~10,000 sequences per selection round due to multiplexing of numerous selection outputs in one chip. This number was sufficient to interrogate the selection output and isolate a few low-frequency rare clones from the selection output. Advances in depth and length of NGS reads will help to obtain more diverse Fab clones and to reconstruct ultra-low frequency Fabs from earlier rounds of selection. This NGS-assisted approach can also be used for screening Fabs against an array of antigens pooled in different configurations, and subsequently deconvoluting the resulting selection outputs to deduce Fab sequences specific for each pooled antigen (Larman *et al*., 2012). NGS information can also be used for mapping binding energy landscapes, designing affinity maturation libraries, and improving biophysical properties of antibodies, especially when crystal structures of antigen-antibody complexes are also available for analysis (Whitehead *et al*., 2012; Koenig *et al*., 2015; Koenig *et al*., 2017).

While generating Notch-1 binders from single-framework synthetic Fab libraries, we realized the potential of screening focused diversity libraries in obtaining Fabs with both high affinity and high specificity. In the case of library-S, screening sub-libraries identified a low-nM Notch-1 specific binder. In the case of library-F, fine-tuning the library towards long CDR lengths gave rise to a sub-nM Notch-1 specific binder. Our work showed that screening focused diversity libraries could provide new and better binding solutions in comparison with Fabs obtained from the master library. Sub-libraries and long-CDR libraries can be used in high-throughput selection campaigns for generating new and better binding solutions against a large panel of antigens. In particular, the long-CDR library can be useful for generating antagonists against enzymes in which the druggable active sites are buried deeply inside the protein (Nam *et al*., 2016). Interestingly, the long-CDR library (Modified-F) gave rise to a low-nM binder against

MERTK, a target for which both the other libraries (S and F) proved to be unsuccessful. The anti-MERTK Fab contained design features found only in modified-F, but not in library-S (CDRL3 length of 12 residues and H1/H2 optimization) or library-F (CDRH3 length of 24 residues and J_H6 usage). This indicated that small changes to diversity designs could increase the success rate of single-framework synthetic Fab libraries.

The new synthetic Fab library, library-S, constructed in this work delivered promising Fabs against Jagged-1, Jagged-2 and Notch-1 and was moderately successful against Notch-2 and Notch-3. The highest-affinity Fab, named J2/S/4, bound to Jagged-2 with a K_D value of 80 ± 13 pM and showed >1000-fold specificity over Jagged-1. J2/S/4 possessed ~3-fold higher affinity than the Jagged-2 specific synthetic antibody anti-JAG2.b33 reported previously (K_D = 210 pM) (Lafkas *et al*., 2015). Library-F, constructed and validated by the Sidhu lab (Persson *et al*., 2013), delivered numerous high affinity Fabs against Jagged-2 and Notch-1. The new modified-F long-CDR library gave rise to the highest affinity Notch-1 binder. The Fab clone, named N1/ModF/2, bound to Notch-1 with a K_D value of 0.15 ± 0.01 nM and exhibited strict specificity for Notch-1. N1/ModF/2 possessed ~17-fold higher affinity than the Notch-1 specific synthetic antibody anti-NRR1 reported previously (K_D = 2.5 nM) (Wu *et al*., 2010). We generated Jagged-2 Fabs with 12 different CDRH3 lengths and Notch-1 Fabs with 10 different CDRH3 lengths. Since these Fabs exhibited different binding profiles, we speculate that these Fabs recognize different epitopes on Jagged-2 and Notch-1 and could be used for modulating the Notch signaling pathway using different mechanisms of action. In contrast to γ-secretase inhibitors and Notch monoclonal antibodies that suffer from treatment-related side effects due to pan-Notch inhibition (Espinoza and Miele, 2013; Takebe *et al*., 2014), our mono-specific Notch-1 and Jagged-2 Fabs possessing clinically relevant K_D values will permit a more precise modulation of Notch signaling pathway for both research and therapeutic applications.

Previous studies have shown that linear peptides, discrete motifs or small domains can be grafted into CDRs to design antibodies with desired binding properties (Simon *et al*., 2005; Perchiacca *et al*., 2012; Zhang *et al*., 2014; Liu *et al*., 2015B). A few studies have combined motif grafting with combinatorial approaches to improve the binding properties of grafted antibodies (Barbas *et al*., 1993; Koerber *et al*., 2013; Lee *et al*., 2015). In this motif-grafting project, we used structural information on the EGFR homo-dimerization interaction to design a single-framework synthetic Fab library that was biased towards interacting with domain II of

EGFR. The structure-guided Fab library was constructed by grafting the EGFR domain-II dimerization arm into CDRH3 and by diversifying key positions within ungrafted CDRs. Screening the structure-guided Fab library against EGFR-ECD gave rise to a Fab named DL06. For comparison, Library-F was used to generate an anti-EGFR Fab named Fab-H whose binding overlapped with the Fab-DL06. Both Fabs possessed low-nM K_D values for recombinant and cell-surface EGFR and inhibited EGF-mediated EGFR activation. Epitopes of both Fabs mapped to EGFR domains I-II and did not overlap with binding sites of any clinically-relevant anti-EGFR antibodies.

The motif-grafting approach described in this work can be used as a general strategy for semi-rationally designing phage-displayed synthetic Fab libraries that are biased towards interacting with disease-relevant discontinuous protein-protein interfaces. If structural information about the antigen is available, this strategy can be extended to target difficult or buried epitopes on other receptors. The EGFR domain II structure-guided Fab library constructed in this project can be used for generating Fabs against other ErbB receptors, and for conducting biopanning experiments against cell-surface EGFR. The EGFR domain truncations and dimerization loop mutants constructed in this project can be used as targets in biopanning experiments for generating Fabs against specific EGFR domains, motifs or conformations. Both anti-EGFR Fabs developed in this project possess low-nM K_D values, recognize new epitopes on EGFR, and inhibit EGF-induced EGFR activation, thus they are good candidates for structural studies and cell biological studies. Since DL06 and FabH exhibited differences in their *in vitro* binding kinetics despite targeting similar epitopes, and considering that DL06 contains an ultra-long CDRH3 with 38 residues, it will be interesting to compare the binding thermodynamics and binding mechanisms of DL06 and FabH. In combination with cetuximab, DL06 and FabH can be used for engineering dual-domain-targeting antibodies against EGFR. Currently, other members of the laboratory are developing positron emission tomography-based methods for detecting EGFR overexpression *in vivo* using DL06/FabH-derived imaging probes.

In total, this PhD project resulted in novel methods for discovering synthetic antibodies, three functional Fab libraries and numerous high-quality Fabs against Notch, Jagged and EGF receptors. We showed that implementing NGS approaches, screening focused diversity libraries, fine-tuning the CDR length diversity, making changes to CDR diversity designs and using structure-guided motif-grafting approaches can improve the success rate of single-framework

synthetic Fab libraries. These findings have valuable implications for antibody library design, antibody phage display and combinatorial antibody engineering. The NGS-guided antibody discovery platform and Fab libraries developed in this work may result in reliable and sustainable isolation of Fabs for multiple disease-relevant antigens. Further, Fabs discovered in this work can serve as lead molecules for developing novel Notch/Jagged/EGFR-targeted therapeutics such as mono/bi-specific antibodies, T-cell redirecting antibodies, antibody-drug conjugates, antibody-radionuclide conjugates and chimeric antigen receptor T cells.

7. REFERENCES

Ackerman, M.E., Lai, J.I., Pastan, I., and Wittrup, K.D. (2011). Exploiting bias in a non-immune human antibody library to predict antigenicity. Protein Eng. Des. Sel. *24*, 845-853.

Adams, J.J., and Sidhu, S.S. (2014). Synthetic antibody technologies. Curr. Opin. Struct. Biol. *24*, 1-9.

Adams, J.J., Nelson, B., and Sidhu, S.S. (2014). Recombinant genetic libraries and human monoclonal antibodies. Methods Mol. Biol. *1060*, 149-170.

Aggarwal, S. (2009). What's fueling the biotech engine-2008? Nat. Biotechnol. *27*, 987-993.

Al-Lazikani, B., Lesk, A.M., and Chothia, C. (1997). Standard conformations for the canonical structures of immunoglobulins. J. Mol. Biol. *273*, 927-948.

Almagro, J.C., and Fransson, J. (2008). Humanization of antibodies. Front. Biosci. *13*, 1619-1633.

Arteaga, C.L., and Engelman, J.A. (2014). ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics. Cancer cell *25*, 282-303.

Ausubel, F.M., Brent, R., Kingston R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1997). Current Protocols in Molecular Biology (New York, USA: John Wiley & Sons).

Bai, X., Kim, J., Kang, S., Kim, W., and Shim, H. (2015). A Novel Human scFv Library with Non-Combinatorial Synthetic CDR Diversity. PLoS One *10*, e0141045.

Barbas, C.F., 3rd, Kang, A.S., Lerner, R.A., and Benkovic, S.J. (1991). Assembly of combinatorial antibody libraries on phage surfaces: the gene III site. Proc. Natl. Acad. Sci. USA *88*, 7978-7982.

Barbas, C.F., 3rd, Languino, L.R., and Smith, J.W. (1993). High-affinity self-reactive human antibodies by design and selection: targeting the integrin ligand binding site. Proc. Natl. Acad. Sci. USA *90*, 10003-10007.

Barthelemy, P.A., Raab, H., Appleton, B.A., Bond, C.J., Wu, P., Wiesmann, C., and Sidhu, S.S. (2008). Comprehensive analysis of the factors contributing to the stability and solubility of autonomous human VH domains. J. Biol. Chem. *283*, 3639-3654.

Beck, A., Wurch, T., Bailly, C., and Corvaia, N. (2010). Strategies and challenges for the next generation of therapeutic antibodies. Nat. Rev. Immunol. *10*, 345-352.

Benhar, I. (2007). Design of synthetic antibody libraries. Expert Opin. Biol. Ther. *7*, 763-779.

Bessman, N.J., Bagchi, A., Ferguson, K.M., and Lemmon, M.A. (2014). Complex relationship between ligand binding and dimerization in the epidermal growth factor receptor. Cell Rep. *9*, 1306-1317.

Birtalan, S., Zhang, Y., Fellouse, F.A., Shao, L., Schaefer, G., and Sidhu, S.S. (2008). The intrinsic contributions of tyrosine, serine, glycine and arginine to the affinity and specificity of antibodies. J. Mol. Biol. *377*, 1518-1528.

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a web-based genome analysis tool for experimentalists. Curr. Protoc. Mol. Biol. Chapter 19, Unit 19, 11-21.

Bradbury, A.R., Sidhu, S., Dubel, S., and McCafferty, J. (2011). Beyond natural antibodies: the power of in vitro display technologies. Nat. Biotechnol. *29*, 245-254.

Burgess, A.W., Cho, H.S., Eigenbrot, C., Ferguson, K.M., Garrett, T.P., Leahy, D.J., Lemmon, M.A., Sliwkowski, M.X., Ward, C.W., and Yokoyama, S. (2003). An open-and-shut case? Recent insights into the activation of EGF/ErbB receptors. Mol. Cell *12*, 541-552.

Burkovitz, A., and Ofran, Y. (2016). Understanding differences between synthetic and natural antibodies can help improve antibody engineering. mAbs *8*, 278-287.

Burton, D.R., Barbas, C.F., 3rd, Persson, M.A., Koenig, S., Chanock, R.M., and Lerner, R.A. (1991). A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic seropositive individuals. Proc. Natl. Acad. Sci. USA *88*, 10134-10137.

Camidge, D.R., Pao, W., and Sequist, L.V. (2014). Acquired resistance to TKIs in solid tumours: learning from lung cancer. Nat. Rev. Clin. Oncol. *11*, 473-481.

Carter, P., Presta, L., Gorman, C.M., Ridgway, J.B., Henner, D., Wong, W.L., Rowland, A.M., Kotts, C., Carver, M.E., and Shepard, H.M. (1992). Humanization of an anti-p185HER2 antibody for human cancer therapy. Proc. Natl. Acad. Sci. USA *89*, 4285-4289.

Carter, R.E., and Sorkin, A. (1998). Endocytosis of functional epidermal growth factor receptor-green fluorescent protein chimera. J. Biol. Chem. *273*, 35000-35007.

Chames, P., Van Regenmortel, M., Weiss, E., and Baty, D. (2009). Therapeutic antibodies: successes, limitations and hopes for the future. Br. J. Pharmacol. *157*, 220-233.

Chan, A.C., and Carter, P.J. (2010). Therapeutic antibodies for autoimmunity and inflammation. Nat. Rev. Immunol. *10*, 301-316.

Chan, C.E., Lim, A.P., MacAry, P.A., and Hanson, B.J. (2014). The role of phage display in therapeutic antibody discovery. Int. Immunol. *26*, 649-657.

Chen, G., and Sidhu, S.S. (2014). Design and generation of synthetic antibody libraries for phage display. Methods Mol. Biol. *1131*, 113-131.

Chiu, M.L., and Gilliland, G.L. (2016). Engineering antibody therapeutics. Curr. Opin. Struct. Biol. *38*, 163-173.

Chothia, C., Lesk, A.M., Gherardi, E., Tomlinson, I.M., Walter, G., Marks, J.D., Llewelyn, M.B., and Winter, G. (1992). Structural repertoire of the human VH segments. J. Mol. Biol. *227*, 799-817.

Clackson, T., Hoogenboom, H.R., Griffiths, A.D., and Winter, G. (1991). Making antibody fragments using phage display libraries. Nature *352*, 624-628.

Clark, M.A., Hawkins, N.J., Papaioannou, A., Fiddes, R.J., and Ward, R.L. (1997). Isolation of human anti-c-erbB-2 Fabs from a lymph node-derived phage display library. Clin. Exp. Immunol. *109*, 166-174.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: a sequence logo generator. Genome Res. *14*, 1188-1190.

D'Angelo, S., Kumar, S., Naranjo, L., Ferrara, F., Kiss, C., and Bradbury, A.R. (2014). From deep sequencing to actual clones. Protein Eng. Des. Sel. *27*, 301-307.

Dawson, J.P., Berger, M.B., Lin, C.C., Schlessinger, J., Lemmon, M.A., and Ferguson, K.M. (2005). Epidermal growth factor receptor dimerization and activation require ligand-induced conformational changes in the dimer interface. Mol. Cell. Biol. *25*, 7734-7742.

DeKosky, B.J., Kojima, T., Rodin, A., Charab, W., Ippolito, G.C., Ellington, A.D., and Georgiou, G. (2015). In-depth determination and analysis of the human paired heavy- and light-chain antibody repertoire. Nat. Med. *21*, 86-91.

Delves, P.J., Martin, S.J., Burton, D.R., and Roitt, I.M. (2017). Roitt's essential immunology, 13th edition (Chichester, UK: Wiley-Blackwell), pp. 1-96.

Dennis, M.S. (2015). Selection and screening strategies. In Phage display in biotechnology and drug discovery, S.S. Sidhu, C.R. Geyer, eds. (Boca Raton, USA: Taylor and Francis), pp. 97-112.

Derda, R., Tang, S.K., Li, S.C., Ng, S., Matochko, W., and Jafari, M.R. (2011). Diversity of phage-displayed libraries of peptides during panning and amplification. Molecules *16*, 1776-1803.

Dübel, S., and Reichert, J.M. (2014). Therapeutic antibodies: from past to future. In Handbook of therapeutic antibodies, S. Dubel, J.M. Reichert, eds. (Weinheim, Germany: Wiley-Blackwell), pp. 1-13.

Ecker, D.M., Jones, S.D., and Levine, H.L. (2015). The therapeutic monoclonal antibody market. mAbs *7*, 9-14.

Eigenbrot, C., Randal, M., Presta, L., Carter, P., and Kossiakoff, A.A. (1993). X-ray structures of the antigen-binding domains from three variants of humanized anti-p185HER2 antibody 4D5 and comparison with molecular modeling. J. Mol. Biol. *229*, 969-995.

Emmons, C., and Hunsicker, L.G. (1987). Muromonab-CD3 (Orthoclone OKT3): the first monoclonal antibody approved for therapeutic use. Iowa Med. *77*, 78-82.

Espinoza, I., and Miele, L. (2013). Notch inhibitors for cancer treatment. Pharmacol. Ther. *139*, 95-110.

Fauvel, B., and Yasri, A. (2014). Antibodies directed against receptor tyrosine kinases: current and future strategies to fight cancer. mAbs *6*, 838-851.

Fellouse, F.A., and Pal, G. (2015). Methods for the construction of phage-displayed libraries. In Phage display in biotechnology and drug discovery, S.S. Sidhu, C.R. Geyer, eds. (Boca Raton, USA: Taylor and Francis), pp. 75-96.

Fellouse, F.A., and Sidhu, S.S. (2005). Synthetic antibody libraries. In Phage display in biotechnology and drug discovery, S.S. Sidhu, ed. (Boca Raton, USA: Taylor and Francis), pp. 709-740.

Fellouse, F.A., and Sidhu, S.S. (2006). Making antibodies in bacteria. In Making and using antibodies: a practical handbook, G.C. Howard, M.R. Kaser, eds. (San Francisco, USA: Taylor and Francis), pp. 157–180.

Fellouse, F.A., and Sidhu, S.S. (2015). Synthetic antibody libraries. In Phage display in biotechnology and drug discovery, S.S. Sidhu, C.R. Geyer, eds. (Boca Raton, USA: Taylor and Francis), pp. 495-520.

Fellouse, F.A., Barthelemy, P.A., Kelley, R.F., and Sidhu, S.S. (2006). Tyrosine plays a dominant functional role in the paratope of a synthetic antibody derived from a four amino acid code. J. Mol. Biol. *357*, 100-114.

Fellouse, F.A., Esaki, K., Birtalan, S., Raptis, D., Cancasci, V.J., Koide, A., Jhurani, P., Vasser, M., Wiesmann, C., Kossiakoff, A.A*., et al.* (2007). High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries. J. Mol. Biol. *373*, 924-940.

Fellouse, F.A., Li, B., Compaan, D.M., Peden, A.A., Hymowitz, S.G., and Sidhu, S.S. (2005). Molecular recognition by a binary code. J. Mol. Biol. *348*, 1153-1162.

Fellouse, F.A., Wiesmann, C., and Sidhu, S.S. (2004). Synthetic antibodies from a four-amino-acid code: a dominant role for tyrosine in antigen recognition. Proc. Natl. Acad. Sci. USA *101*, 12467-12472.

Ferguson, K.M. (2008). Structure-based view of epidermal growth factor receptor regulation. Annu. Rev. Biophys. *37*, 353-373.

Ferrara, F., Naranjo, L.A., Kumar, S., Gaiotto, T., Mukundan, H., Swanson, B., and Bradbury, A.R. (2012). Using phage and yeast display to select hundreds of monoclonal antibodies: application to antigen 85, a tuberculosis biomarker. PLoS One *7*, e49535.

Finlay, W.J., and Almagro, J.C. (2012). Natural and man-made V-gene repertoires for antibody discovery. Front. Immunol. *3*, 342.

Fischer, N. (2011). Sequencing antibody repertoires: the next generation. mAbs *3*, 17-20.

Fisher, R.D., Ultsch, M., Lingel, A., Schaefer, G., Shao, L., Birtalan, S., Sidhu, S.S., and Eigenbrot, C. (2010). Structure of the complex between HER2 and an antibody paratope formed by side chains from tryptophan and serine. J. Mol. Biol. *402*, 217-229.

Frederickson, S., Renshaw, M.W., Lin, B., Smith, L.M., Calveley, P., Springhorn, J.P., Johnson, K., Wang, Y., Su, X., Shen, Y.*, et al.* (2006). A rationally designed agonist antibody fragment that functionally mimics thrombopoietin. Proc. Natl. Acad. Sci. USA *103*, 14307-14312.

Freise, A.C., and Wu, A.M. (2015). In vivo imaging with antibodies and engineered fragments. Mol. Immunol. *67*, 142-152.

Frenzel, A., Schirrmann, T., and Hust, M. (2016). Phage display-derived human antibodies in clinical development and therapy. mAbs *8*, 1177-1194.

Friedman, L.M., Rinon, A., Schechter, B., Lyass, L., Lavi, S., Bacus, S.S., Sela, M., and Yarden, Y. (2005). Synergistic down-regulation of receptor tyrosine kinases by combinations of mAbs: implications for cancer immunotherapy. Proc. Natl. Acad. Sci. USA *102*, 1915-1920.

Fuh, G., and Sidhu, S.S. (2000). Efficient phage display of polypeptides fused to the carboxy-terminus of the M13 gene-3 minor coat protein. FEBS Lett. *480*, 231-234.

Gan, H.K., Burgess, A.W., Clayton, A.H., and Scott, A.M. (2012). Targeting of a conformationally exposed, tumor-specific epitope of EGFR as a strategy for cancer therapy. Cancer Res. *72*, 2924-2930.

Gan, H.K., Kaye, A.H., and Luwor, R.B. (2009). The EGFRvIII variant in glioblastoma multiforme. J. Clin. Neurosci. *16*, 748-754.

Garrett, T.P., Burgess, A.W., Gan, H.K., Luwor, R.B., Cartwright, G., Walker, F., Orchard, S.G., Clayton, A.H., Nice, E.C., Rothacker, J.*, et al.* (2009). Antibodies specifically targeting a locally misfolded region of tumor associated EGFR. Proc. Natl. Acad. Sci. USA *106*, 5082-5087.

Garrett, T.P., McKern, N.M., Lou, M., Elleman, T.C., Adams, T.E., Lovrecz, G.O., Zhu, H.J., Walker, F., Frenkel, M.J., Hoyne, P.A.*, et al.* (2002). Crystal structure of a truncated epidermal growth factor receptor extracellular domain bound to transforming growth factor alpha. Cell *110*, 763-773.

Georgiou, G., Ippolito, G.C., Beausang, J., Busse, C.E., Wardemann, H., and Quake, S.R. (2014). The promise and challenge of high-throughput sequencing of the antibody repertoire. Nat. Biotechnol. *32*, 158-168.

Geyer, C.R., McCafferty, J., Dubel, S., Bradbury, A.R., and Sidhu, S.S. (2012). Recombinant antibodies and in vitro selection technologies. Methods Mol. Biol. *901*, 11-32.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343-345.

Gilbreth, R.N., Esaki, K., Koide, A., Sidhu, S.S., and Koide, S. (2008). A dominant conformational role for amino acid diversity in minimalist protein-protein interfaces. J. Mol. Biol. *381*, 407-418.

Gitlin, A.D., Shulman, Z., and Nussenzweig, M.C. (2014). Clonal selection in the germinal centre by regulated proliferation and hypermutation. Nature *509*, 637-640.

Glanville, J., D'Angelo, S., Khan, T.A., Reddy, S.T., Naranjo, L., Ferrara, F., and Bradbury, A.R. (2015). Deep sequencing in library selection projects: what insight does it bring? Curr. Opin. Struct. Biol. *33*, 146-160.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. *11*, R86.

Gordon, W.R., Arnett, K.L., and Blacklow, S.C. (2008). The molecular logic of Notch signaling-a structural and biochemical perspective. J. Cell Sci. *121*, 3109-3119.

Graus, Y.F., de Baets, M.H., Parren, P.W., Berrih-Aknin, S., Wokke, J., van Breda Vriesman, P.J., and Burton, D.R. (1997). Human anti-nicotinic acetylcholine receptor recombinant Fab fragments isolated from thymus-derived phage display libraries from myasthenia gravis patients reflect predominant specificities in serum and block the action of pathogenic serum antibodies. J. Immunol. *158*, 1919-1929.

Groves, M., Lane, S., Douthwaite, J., Lowne, D., Rees, D.G., Edwards, B., and Jackson, R.H. (2006). Affinity maturation of phage display antibody populations using ribosome display. J. Immunol. Methods *313*, 129-139.

Harel-Inbar, N., and Benhar, I. (2012). Selection of antibodies from synthetic antibody libraries. Arch. Biochem. Biophys. *526*, 87-98.

Held, H.A., and Sidhu, S.S. (2004). Comprehensive mutational analysis of the M13 major coat protein: improved scaffolds for C-terminal phage display. J. Mol. Biol. *340*, 587-597.

Hoet, R.M., Cohen, E.H., Kent, R.B., Rookey, K., Schoonbroodt, S., Hogan, S., Rem, L., Frans, N., Daukandt, M., Pieters, H.*, et al.* (2005). Generation of high-affinity human antibodies by combining donor-derived and synthetic complementarity-determining-region diversity. Nat. Biotechnol. *23*, 344-348.

Hoogenboom, H.R. (2005). Selecting and screening recombinant antibody libraries. Nat. Biotechnol. *23*, 1105-1116.

Hornsby, M., Paduch, M., Miersch, S., Saaf, A., Matsuguchi, T., Lee, B., Wypisniak, K., Doak, A., King, D., Usatyuk, S.*, et al.* (2015). A High Through-put Platform for Recombinant Antibodies to Folded Proteins. Mol. Cell Proteomics *14*, 2833-2847.

Huang, M., Anand, S., Murphy, E.A., Desgrosellier, J.S., Stupack, D.G., Shattil, S.J., Schlaepfer, D.D., and Cheresh, D.A. (2012). EGFR-dependent pancreatic carcinoma cell metastasis through Rap1 activation. Oncogene *31*, 2783-2793.

Huang, Y., Bharill, S., Karandur, D., Peterson, S.M., Marita, M., Shi, X., Kaliszewski, M.J., Smith, A.W., Isacoff, E.Y., and Kuriyan, J. (2016). Molecular basis for multimerization in the activation of the epidermal growth factor receptor. ELife *5*.

Hudziak, R.M., Schlessinger, J., and Ullrich, A. (1987). Increased expression of the putative growth factor receptor p185HER2 causes transformation and tumorigenesis of NIH 3T3 cells. Proc. Natl. Acad. Sci. USA *84*, 7159-7163.

Huse, W.D., Sastry, L., Iverson, S.A., Kang, A.S., Alting-Mees, M., Burton, D.R., Benkovic, S.J., and Lerner, R.A. (1989). Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda. Science *246*, 1275-1281.

Jara-Acevedo, R., Diez, P., Gonzalez-Gonzalez, M., Degano, R.M., Ibarrola, N., Gongora, R., Orfao, A., and Fuentes, M. (2016). Methods for selecting phage display antibody libraries. Curr. Pharm. Des. *22*, 6490-6499.

Jubb, H., Higueruelo, A.P., Winter, A., and Blundell, T.L. (2012). Structural biology and drug discovery for protein-protein interactions. Trends Pharmacol. Sci. *33*, 241-248.

Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S., and Foeller, C. (1991). Sequences of proteins of immunological interest. 5th ed. (Bethesda, USA: National Institutes of Health).

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C.*, et al.* (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics *28*, 1647-1649.

Kelley, R.F., O'Connell, M.P., Carter, P., Presta, L., Eigenbrot, C., Covarrubias, M., Snedecor, B., Bourell, J.H., and Vetterlein, D. (1992). Antigen binding thermodynamics and antiproliferative effects of chimeric and humanized anti-p185HER2 antibody Fab fragments. Biochemistry *31*, 5434-5441.

Klapper, L.N., Vaisman, N., Hurwitz, E., Pinkas-Kramarski, R., Yarden, Y., and Sela, M. (1997). A subclass of tumor-inhibitory monoclonal antibodies to ErbB-2/HER2 blocks crosstalk with growth factor receptors. Oncogene *14*, 2099-2109.

Knappik, A., Ge, L., Honegger, A., Pack, P., Fischer, M., Wellnhofer, G., Hoess, A., Wolle, J., Pluckthun, A., and Virnekas, B. (2000). Fully synthetic human combinatorial antibody libraries (HuCAL) based on modular consensus frameworks and CDRs randomized with trinucleotides. J. Mol. Biol. *296*, 57-86.

Ko, B.K., Lee, S.Y., Lee, Y.H., Hwang, I.S., Persson, H., Rockberg, J., Borrebaeck, C., Park, D., Kim, K.T., Uhlen, M.*, et al.* (2015). Combination of novel HER2-targeting antibody 1E11 with trastuzumab shows synergistic antitumor activity in HER2-positive gastric cancer. Mol. Oncol. *9*, 398-408.

Koenig, P., Lee, C.V., Sanowar, S., Wu, P., Stinson, J., Harris, S.F., and Fuh, G. (2015). Deep Sequencing-guided Design of a High Affinity Dual Specificity Antibody to Target Two

Angiogenic Factors in Neovascular Age-related Macular Degeneration. J. Biol. Chem. *290*, 21773-21786.

Koenig, P., Lee, C.V., Walters, B.T., Janakiraman, V., Stinson, J., Patapoff, T.W., and Fuh, G. (2017). Mutational landscape of antibody variable domains reveals a switch modulating the interdomain conformational dynamics and antigen binding. Proc. Natl. Acad. Sci. USA *114*, E486-E495.

Koerber, J.T., Thomsen, N.D., Hannigan, B.T., Degrado, W.F., and Wells, J.A. (2013). Nature-inspired design of motif-specific antibody scaffolds. Nat. Biotechnol. *31*, 916-921.

Kogelberg, H., Tolner, B., Thomas, G.J., Di Cara, D., Minogue, S., Ramesh, B., Sodha, S., Marsh, D., Lowdell, M.W., Meyer, T.*, et al.* (2008). Engineering a single-chain Fv antibody to alpha v beta 6 integrin using the specificity-determining loop of a foot-and-mouth disease virus. J. Mol. Biol. *382*, 385-401.

Kohler, G., and Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. Nature *256*, 495-497.

Koide, A., Wojcik, J., Gilbreth, R.N., Reichel, A., Piehler, J., and Koide, S. (2009). Accelerating phage-display library selection by reversible and site-specific biotinylation. Protein Eng. Des. Sel. *22*, 685-690.

Koide, S., and Sidhu, S.S. (2009). The importance of being tyrosine: lessons in molecular recognition from minimalist synthetic binding proteins. ACS Chem. Biol. *4*, 325-334.

Kruser, T.J., and Wheeler, D.L. (2010). Mechanisms of resistance to HER family targeting antibodies. Exp. Cell Res. *316*, 1083-1100.

Kugler, J., Wilke, S., Meier, D., Tomszak, F., Frenzel, A., Schirrmann, T., Dubel, S., Garritsen, H., Hock, B., Toleikis, L.*, et al.* (2015). Generation and analysis of the improved human HAL9/10 antibody phage display libraries. BMC Biotechnol. *15*, 10.

Kunkel, T.A., Roberts, J.D., and Zakour, R.A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. *154*, 367-382.

Ladiwala, A.R., Bhattacharya, M., Perchiacca, J.M., Cao, P., Raleigh, D.P., Abedini, A., Schmidt, A.M., Varkey, J., Langen, R., and Tessier, P.M. (2012). Rational design of potent domain antibody inhibitors of amyloid fibril assembly. Proc. Natl. Acad. Sci. USA *109*, 19965-19970.

Lafkas, D., Shelton, A., Chiu, C., de Leon Boenig, G., Chen, Y., Stawicki, S.S., Siltanen, C., Reichelt, M., Zhou, M., Wu, X.*, et al.* (2015). Therapeutic antibodies reveal Notch control of transdifferentiation in the adult lung. Nature *528*, 127-131.

Larman, H.B., Xu, G.J., Pavlova, N.N., and Elledge, S.J. (2012). Construction of a rationally designed antibody platform for sequencing-assisted selection. Proc. Natl. Acad. Sci. USA *109*, 18523-18528.

Lavinder, J.J., Horton, A.P., Georgiou, G., and Ippolito, G.C. (2015). Next-generation sequencing and protein mass spectrometry for the comprehensive analysis of human cellular and serum antibody repertoires. Curr. Opin. Chem. Biol. *24*, 112-120.

Lee, C.C., Julian, M.C., Tiller, K.E., Meng, F., DuConge, S.E., Akter, R., Raleigh, D.P., and Tessier, P.M. (2016). Design and Optimization of Anti-amyloid Domain Antibodies Specific for beta-Amyloid and Islet Amyloid Polypeptide. J. Biol. Chem. *291*, 2858-2873.

Lee, C.V., Liang, W.C., Dennis, M.S., Eigenbrot, C., Sidhu, S.S., and Fuh, G. (2004B). High-affinity human antibodies from phage-displayed synthetic Fab libraries with a single framework scaffold. J. Mol. Biol. *340*, 1073-1093.

Lee, C.V., Sidhu, S.S., and Fuh, G. (2004A). Bivalent antibody phage display mimics natural immunoglobulin. J. Immunol. Methods *284*, 119-132.

Lemmon, M.A., Schlessinger, J., and Ferguson, K.M. (2014). The EGFR family: not so prototypical receptor tyrosine kinases. Cold Spring Harb. Perspect. Biol. *6*, a020768.

Liu, T., Fu, G., Luo, X., Liu, Y., Wang, Y., Wang, R.E., Schultz, P.G., and Wang, F. (2015B). Rational design of antibody protease inhibitors. J. Am. Chem. Soc. *137*, 4042-4045.

Liu, T., Zhang, Y., Liu, Y., Wang, Y., Jia, H., Kang, M., Luo, X., Caballero, D., Gonzalez, J., Sherwood, L.*, et al.* (2015A). Functional human antibody CDR fusions as long-acting therapeutic endocrine agonists. Proc. Natl. Acad. Sci. USA *112*, 1356-1361.

Liu, Y., Wang, Y., Zhang, Y., Liu, T., Jia, H., Zou, H., Fu, Q., Zhang, Y., Lu, L., Chao, E.*, et al.* (2016). Rational Design of Dual Agonist-Antibody Fusions as Long-acting Therapeutic Hormones. ACS Chem. Biol. *11*, 2991-2995.

Lloyd, C., Lowe, D., Edwards, B., Welsh, F., Dilks, T., Hardman, C., and Vaughan, T. (2009). Modelling the human immune response: performance of a 1011 human antibody repertoire against a broad panel of therapeutically relevant antigens. Protein Eng. Des. Sel. *22*, 159-168.

Loman, N.J., Misra, R.V., Dallman, T.J., Constantinidou, C., Gharbia, S.E., Wain, J., and Pallen, M.J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. Nat. Biotechnol. *30*, 434-439.

Lonberg, N. (2008). Human monoclonal antibodies from transgenic mice. Handb. Exp. Pharmacol. 69-97.

Lopez, T., Nam, D.H., Kaihara, E., Mustafa, Z., and Ge, X. (2017). Identification of highly selective MMP-14 inhibitory Fabs by deep sequencing. Biotechnol. Bioeng. *114*, 1140-1150.

Lovgren, J., Pursiheimo, J.P., Pyykko, M., Salmi, J., and Lamminmaki, U. (2016). Next generation sequencing of all variable loops of synthetic single framework scFv-Application in anti-HDL antibody selections. N. Biotechnol. *33*, 790-796.

Ma, X., Barthelemy, P.A., Rouge, L., Wiesmann, C., and Sidhu, S.S. (2013). Design of synthetic autonomous VH domain libraries and structural analysis of a VH domain bound to vascular endothelial growth factor. J. Mol. Biol. *425*, 2247-2259.

Mahon, C.M., Lambert, M.A., Glanville, J., Wade, J.M., Fennell, B.J., Krebs, M.R., Armellino, D., Yang, S., Liu, X., O'Sullivan, C.M., *et al.* (2013). Comprehensive interrogation of a minimalist synthetic CDR-H3 library and its ability to generate antibodies with therapeutic potential. J. Mol. Biol. *425*, 1712-1730.

Male, D., Brostoff, J., Roth, D.B., and Roitt, I.M. (2006). Immunology, 7th international edition (Maryland Heights, USA: Mosby-Elsevier), pp.3-86.

Maruyama, T., Rodriguez, L.L., Jahrling, P.B., Sanchez, A., Khan, A.S., Nichol, S.T., Peters, C.J., Parren, P.W., and Burton, D.R. (1999). Ebola virus can be effectively neutralized by antibody produced in natural human infection. J. Virol. *73*, 6024-6030.

Marvin, J.S., and Lowman, H.B. (2015). Antibody humanization and affinity maturation using phage display. In Phage display in biotechnology and drug discovery, S.S. Sidhu, C.R. Geyer, eds. (Boca Raton, USA: Taylor and Francis), pp. 347-372.

Mathonet, P., and Ullman, C.G. (2013). The application of next generation sequencing to the understanding of antibody repertoires. Front. Immunol. *4*, 265.

McCafferty, J., Griffiths, A.D., Winter, G., and Chiswell, D.J. (1990). Phage antibodies: filamentous phage displaying antibody variable domains. Nature *348*, 552-554.

McLane, K.E., Burton, D.R., and Ghazal, P. (1995). Transplantation of a 17-amino acid alpha-helical DNA-binding domain into an antibody molecule confers sequence-dependent DNA recognition. Proc. Natl. Acad. Sci. USA *92*, 5214-5218.

Metzker, M.L. (2010). Sequencing technologies- the next generation. Nat. Rev. Genet. *11*, 31-46.

Michnick, S.W., and Sidhu, S.S. (2008). Submitting antibodies to binding arbitration. Nat.Chem. Biol. *4*, 326-329.

Miersch, S., and Sidhu, S.S. (2012). Synthetic antibodies: concepts, potential and practical considerations. Methods *57*, 486-498.

Miersch, S., Maruthachalam, B.V., Geyer, C.R., and Sidhu, S.S. (2017). Structure-directed and tailored diversity synthetic antibody libraries yield novel anti-EGFR antagonists. ACS Chem. Biol.

Moldenhauer, G. (2014). Selection strategies for monoclonal antibodies. In Handbook of therapeutic antibodies, S. Dubel, J.M. Reichert, eds. (Weinheim, Germany: Wiley-Blackwell), pp. 17-41.

Montagut, C., Dalmases, A., Bellosillo, B., Crespo, M., Pairet, S., Iglesias, M., Salido, M., Gallen, M., Marsters, S., Tsai, S.P., *et al.* (2012). Identification of a mutation in the extracellular domain of the Epidermal Growth Factor Receptor conferring cetuximab resistance in colorectal cancer. Nat. Med. *18*, 221-223.

Moroncini, G., Kanu, N., Solforosi, L., Abalos, G., Telling, G.C., Head, M., Ironside, J., Brockes, J.P., Burton, D.R., and Williamson, R.A. (2004). Motif-grafted antibodies containing

the replicative interface of cellular PrP are specific for PrPSc. Proc. Natl. Acad. Sci. USA *101*, 10404-10409.

Na, H., Laver, J.D., Jeon, J., Singh, F., Ancevicius, K., Fan, Y., Cao, W.X., Nie, K., Yang, Z., Luo, H.*, et al.* (2016). A high-throughput pipeline for the production of synthetic antibodies for analysis of ribonucleoprotein complexes. RNA *22*, 636-655.

Nam, D.H., Rodriguez, C., Remacle, A.G., Strongin, A.Y., and Ge, X. (2016). Active-site MMP-selective antibody inhibitors discovered from convex paratope synthetic libraries. Proc. Natl. Acad. Sci. USA *113*, 14970-14975.

Naso, M.F., Lu, J., and Panavas, T. (2014). Deep sequencing approaches to antibody discovery. Curr. Drug Discov. Technol. *11*, 85-95.

Nelson, A.L., Dhimolea, E., and Reichert, J.M. (2010). Development trends for human monoclonal antibody therapeutics. Nat. Rev. Drug Discov. *9*, 767-774.

Nelson, B., and Sidhu, S.S. (2012). Synthetic antibody libraries. Methods Mol. Biol. *899*, 27-41.

Nero, T.L., Morton, C.J., Holien, J.K., Wielens, J., and Parker, M.W. (2014). Oncogenic protein interfaces: small molecules, big challenges. Nat. Rev. Cancer *14*, 248-262.

Nixon, A.E., Sexton, D.J., and Ladner, R.C. (2014). Drugs derived from phage display: from candidate identification to clinical practice. mAbs *6*, 73-85.

North, B., Lehmann, A., and Dunbrack, R.L., Jr. (2011). A new clustering of antibody CDR loop conformations. J. Mol. Biol. *406*, 228-256.

Ogiso, H., Ishitani, R., Nureki, O., Fukai, S., Yamanaka, M., Kim, J.H., Saito, K., Sakamoto, A., Inoue, M., Shirouzu, M.*, et al.* (2002). Crystal structure of the complex of human epidermal growth factor and receptor extracellular domains. Cell *110*, 775-787.

Paduch, M., Koide, A., Uysal, S., Rizk, S.S., Koide, S., and Kossiakoff, A.A. (2013). Generating conformation-specific synthetic antibodies to trap proteins in selected functional states. Methods *60*, 3-14.

Pal, G., Kouadio, J.L., Artis, D.R., Kossiakoff, A.A., and Sidhu, S.S. (2006). Comprehensive and quantitative mapping of energy landscapes for protein-protein interactions by rapid combinatorial scanning. J. Biol. Chem. *281*, 22378-22385.

Peipp, M., Dechant, M., and Valerius, T. (2008). Effector mechanisms of therapeutic antibodies against ErbB receptors. Curr. Opin. Immunol. *20*, 436-443.

Pendley, C., Schantz, A., and Wagner, C. (2003). Immunogenicity of therapeutic monoclonal antibodies. Curr. Opin. Mol. Ther. *5*, 172-179.

Peng, Y., Zeng, W., Ye, H., Han, K.H., Dharmarajan, V., Novick, S., Wilson, I.A., Griffin, P.R., Friedman, J.M., and Lerner, R.A. (2015). A General Method for Insertion of Functional Proteins within Proteins via Combinatorial Selection of Permissive Junctions. Chem. Biol. *22*, 1134-1143.

Perchiacca, J.M., Ladiwala, A.R., Bhattacharya, M., and Tessier, P.M. (2012). Structure-based design of conformation- and sequence-specific antibodies against amyloid beta. Proc. Natl. Acad. Sci. USA *109*, 84-89.

Persson, H., Ye, W., Wernimont, A., Adams, J.J., Koide, A., Koide, S., Lam, R., and Sidhu, S.S. (2013). CDR-H3 diversity is not required for antigen recognition by synthetic antibodies. J. Mol. Biol. *425*, 803-811.

Ponsel, D., Neugebauer, J., Ladetzki-Baehs, K., and Tissot, K. (2011). High affinity, developability and functional size: the holy grail of combinatorial antibody library generation. Molecules *16*, 3675-3700.

Prassler, J., Thiel, S., Pracht, C., Polzer, A., Peters, S., Bauer, M., Norenberg, S., Stark, Y., Kolln, J., Popp, A.*, et al.* (2011). HuCAL PLATINUM, a synthetic Fab library optimized for sequence diversity and superior performance in mammalian expression systems. J. Mol. Biol. *413*, 261-278.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Satistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Rajan, S., and Sidhu, S.S. (2012). Simplified synthetic antibody libraries. Methods Enzymol. *502*, 3-23.

Ramaraj, T., Angel, T., Dratz, E.A., Jesaitis, A.J., and Mumey, B. (2012). Antigen-antibody interface properties: composition, residue interactions, and features of 53 non-redundant structures. Biochim. Biophys. Acta *1824*, 520-532.

Ranganathan, P., Weaver, K.L., and Capobianco, A.J. (2011). Notch signalling in solid tumours: a little bit of everything but not all the time. Nat. Rev. Cancer *11*, 338-351.

Ravn, U., Didelot, G., Venet, S., Ng, K.T., Gueneau, F., Rousseau, F., Calloud, S., Kosco-Vilbois, M., and Fischer, N. (2013). Deep sequencing of phage display libraries to support antibody discovery. Methods *60*, 99-110.

Ravn, U., Gueneau, F., Baerlocher, L., Osteras, M., Desmurs, M., Malinge, P., Magistrelli, G., Farinelli, L., Kosco-Vilbois, M.H., and Fischer, N. (2010). By-passing in vitro screening--next generation sequencing technologies applied to antibody display and in silico candidate selection. Nucleic Acids Res. *38*, e193.

Reichert, J.M. (2017). Antibodies to watch in 2017. mAbs *9*, 167-181.

Reshetnyak, A.V., Nelson, B., Shi, X., Boggon, T.J., Pavlenco, A., Mandel-Bausch, E.M., Tome, F., Suzuki, Y., Sidhu, S.S., Lax, I.*, et al.* (2013). Structural basis for KIT receptor tyrosine kinase inhibition by antibodies targeting the D4 membrane-proximal region. Proc. Natl. Acad. Sci. USA *110*, 17832-17837.

Robinson, W.H. (2015). Sequencing the functional antibody repertoire--diagnostic and therapeutic discovery. Nat. Rev. Rheumatol. *11*, 171-182.

Rodi, D.J., Mandava, S., and Makowski, L. (2005). Filamentous bacteriophage structure and biology. In Phage display in biotechnology and drug discovery, S.S. Sidhu, ed. (Boca Raton, USA: Taylor and Francis), pp. 1-61.

Rodi, D.J., Mandava, S., and Makowski, L. (2015). Filamentous bacteriophage structure and biology. In Phage display in biotechnology and drug discovery, S.S. Sidhu, C.R. Geyer, eds. (Boca Raton, USA: Taylor and Francis), pp. 1-43.

Roth, T.A., Weiss, G.A., Eigenbrot, C., and Sidhu, S.S. (2002). A minimized M13 coat protein defines the requirements for assembly into the bacteriophage particle. J. Mol. Biol. *322*, 357-367.

Rothberg, J.M., Hinz, W., Rearick, T.M., Schultz, J., Mileski, W., Davey, M., Leamon, J.H., Johnson, K., Milgrew, M.J., Edwards, M.*, et al.* (2011). An integrated semiconductor device enabling non-optical genome sequencing. Nature *475*, 348-352.

Rothe, C., Urlinger, S., Lohning, C., Prassler, J., Stark, Y., Jager, U., Hubner, B., Bardroff, M., Pradel, I., Boss, M.*, et al.* (2008). The human combinatorial antibody library HuCAL GOLD combines diversification of all six CDRs according to the natural immune system with a novel display method for efficient selection of high-affinity antibodies. J. Mol. Biol. *376*, 1182-1200.

Russel, M., Lowman, H.B., and Clackson, T. (2004). Introduction to phage biology and phage display. In Phage display: a practical approach, T. Clackson, H.B. Lowman, eds. (New York, USA: Oxford University Press), pp. 1-24.

Saggy, I., Wine, Y., Shefet-Carasso, L., Nahary, L., Georgiou, G., and Benhar, I. (2012). Antibody isolation from immunized animals: comparison of phage display and antibody discovery via V gene repertoire mining. Protein Eng. Des. Sel. *25*, 539-549.

Sasso, E., Paciello, R., D'Auria, F., Riccio, G., Froechlich, G., Cortese, R., Nicosia, A., De Lorenzo, C., and Zambrano, N. (2015). One-Step Recovery of scFv Clones from High-Throughput Sequencing-Based Screening of Phage Display Libraries Challenged to Cells Expressing Native Claudin-1. BioMed Res. Int. *2015*, 703213.

Schaefer, G., Haber, L., Crocker, L.M., Shia, S., Shao, L., Dowbenko, D., Totpal, K., Wong, A., Lee, C.V., Stawicki, S.*, et al.* (2011). A two-in-one antibody against HER3 and EGFR has superior inhibitory activity compared with monospecific antibodies. Cancer cell *20*, 472-486.

Schlessinger, J. (2002). Ligand-induced, receptor-mediated dimerization and activation of EGF receptor. Cell *110*, 669-672.

Schmidt, E., Hennig, K., Mengede, C., Zillikens, D., and Kromminga, A. (2009). Immunogenicity of rituximab in patients with severe pemphigus. Clin. Immunol. *132*, 334-341.

Schmitz, K.R., and Ferguson, K.M. (2009). Interaction of antibodies with ErbB receptor extracellular regions. Exp. Cell Res. *315*, 659-670.

Scott, A.M., Lee, F.T., Tebbutt, N., Herbertson, R., Gill, S.S., Liu, Z., Skrinos, E., Murone, C., Saunder, T.H., Chappell, B.*, et al.* (2007). A phase I clinical trial with monoclonal antibody

ch806 targeting transitional state and mutant epidermal growth factor receptors. Proc. Natl. Acad. Sci. USA *104*, 4071-4076.

Scott, A.M., Wolchok, J.D., and Old, L.J. (2012). Antibody therapy of cancer. Nat. Rev. Cancer *12*, 278-287.

Scott, D.E., Bayly, A.R., Abell, C., and Skidmore, J. (2016). Small molecules, big targets: drug discovery faces the protein-protein interaction challenge. Nat. Rev. Drug Discov. *15*, 533-550.

Shih, H.H. (2012). Discovery process for antibody-based therapeutics. In Development of antibody-based therapeutics: translational considerations, M.A. Tabrizi, G.G. Bornstein, S.L. Klakamp, eds. (New York, USA: Springer-Verlag), pp. 9-32.

Shim, H. (2015). Synthetic approach to the generation of antibody diversity. BMB Rep. *48*, 489-494.

Shim, H. (2016). Therapeutic Antibodies by Phage Display. Curr. Pharm. Des. *22*, 6538-6559.

Sidhu, S.S. (2012). Antibodies for all: The case for genome-wide affinity reagents. FEBS Lett. *586*, 2778-2779.

Sidhu, S.S., and Fellouse, F.A. (2006). Synthetic therapeutic antibodies. Nat. Chem. Biol. *2*, 682-688.

Sidhu, S.S., and Koide, S. (2007). Phage display for engineering and analyzing protein interaction interfaces. Curr. Opin. Struct. Biol. *17*, 481-487.

Sidhu, S.S., and Kossiakoff, A.A. (2007). Exploring and designing protein function with restricted diversity. Curr. Opin. Chem. Biol. *11*, 347-354.

Sidhu, S.S., and Weiss, G.A. (2004). Constructing phage display libraries by oligonucleotide-directed mutagenesis. In Phage display: a practical approach, T. Clackson, H.B. Lowman, eds. (New York, USA: Oxford University Press), pp. 27-41.

Sidhu, S.S., Feld, B.K., and Weiss, G.A. (2007). M13 bacteriophage coat proteins engineered for improved phage display. Methods Mol. Biol. *352*, 205-219.

Sidhu, S.S., Li, B., Chen, Y., Fellouse, F.A., Eigenbrot, C., and Fuh, G. (2004). Phage-displayed antibody libraries of synthetic heavy chain CDRs. J. Mol. Biol. *338*, 299-310.

Sidhu, S.S., Lowman, H.B., Cunningham, B.C., and Wells, J.A. (2000B). Phage display for selection of novel binding peptides. Methods Enzymol. *328*, 333-363.

Sidhu, S.S., Weiss, G.A., and Wells, J.A. (2000A). High copy display of large proteins on phage for functional selections. J. Mol. Biol. *296*, 487-495.

Simon, P.J., Brogle, K.C., Wang, B., Kyle, D.J., and Soltis, D.A. (2005). Display of somatostatin-related peptides in the complementarity determining regions of an antibody light chain. Arch. Biochem. Biophys. *440*, 148-157.

Smith, G.P. (1985). Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science *228*, 1315-1317.

Soderlind, E., Strandberg, L., Jirholt, P., Kobayashi, N., Alexeiva, V., Aberg, A.M., Nilsson, A., Jansson, B., Ohlin, M., Wingren, C.*, et al.* (2000). Recombining germline-derived CDR sequences for creating diverse single-framework antibody libraries. Nat. Biotechnol. *18*, 852-856.

Spiliotopoulos, A., Owen, J.P., Maddison, B.C., Dreveny, I., Rees, H.C., and Gough, K.C. (2015). Sensitive recovery of recombinant antibody clones after their in silico identification within NGS datasets. J. Immunol. Methods *420*, 50-55.

Sundberg, E.J. (2009). Structural basis of antibody-antigen interactions. Methods Mol. Biol. *524*, 23-36.

Takebe, N., Nguyen, D., and Yang, S.X. (2014). Targeting notch signaling pathway in cancer: clinical development advances and challenges. Pharmacol. Ther.*141*, 140-149.

Tebbutt, N., Pedersen, M.W., and Johns, T.G. (2013). Targeting the ERBB family in cancer: couples therapy. Nat. Rev. Cancer *13*, 663-673.

Tiller, K.E., and Tessier, P.M. (2015). Advances in Antibody Design. Annu. Rev. Biomed. Eng. *17*, 191-216.

Tiller, T., Schuster, I., Deppe, D., Siegers, K., Strohner, R., Herrmann, T., Berenguer, M., Poujol, D., Stehle, J., Stark, Y.*, et al.* (2013). A fully synthetic human Fab antibody library based on fixed VH/VL framework pairings with favorable biophysical properties. mAbs. *5*, 445-470.

Tomic, J., McLaughlin, M., Hart, T., Sidhu, S.S., and Moffat, J. (2015). Leveraging Synthetic Phage-Antibody Libraries for Panning on the Mammalian Cell Surface. In Phage display in biotechnology and drug discovery, S.S. Sidhu, C.R. Geyer, eds. (Boca Raton, USA: Taylor and Francis), pp. 113-122.

Tonikian, R., Zhang, Y., Boone, C., and Sidhu, S.S. (2007). Identifying specificity profiles for peptide recognition modules from phage-displayed peptide libraries. Nat. Protoc. *2*, 1368-1386.

Tundidor, Y., Garcia-Hernandez, C.P., Pupo, A., Cabrera Infante, Y., and Rojas, G. (2014). Delineating the functional map of the interaction between nimotuzumab and the epidermal growth factor receptor. mAbs *6*, 1013-1025.

Vajdos, F.F., Adams, C.W., Breece, T.N., Presta, L.G., de Vos, A.M., and Sidhu, S.S. (2002). Comprehensive functional maps of the antigen-binding site of an anti-ErbB2 antibody obtained with shotgun scanning mutagenesis. J. Mol. Biol. *320*, 415-428.

van den Beucken, T., van Neer, N., Sablon, E., Desmet, J., Celis, L., Hoogenboom, H.R., and Hufton, S.E. (2001). Building novel binding ligands to B7.1 and B7.2 based on human antibody single variable light chain domains. J. Mol. Biol. *310*, 591-601.

Van Regenmortel, M.H. (2009). What is a B-cell epitope? Methods Mol. Biol. *524*, 3-20.

Vaughan, T.J., Williams, A.J., Pritchard, K., Osbourn, J.K., Pope, A.R., Earnshaw, J.C., McCafferty, J., Hodits, R.A., Wilton, J., and Johnson, K.S. (1996). Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. Nat. Biotechnol. *14*, 309-314.

Velu, T.J., Beguinot, L., Vass, W.C., Willingham, M.C., Merlino, G.T., Pastan, I., and Lowy, D.R. (1987). Epidermal-growth-factor-dependent transformation by a human EGF receptor proto-oncogene. Science *238*, 1408-1410.

Venet, S., Ravn, U., Buatois, V., Gueneau, F., Calloud, S., Kosco-Vilbois, M., and Fischer, N. (2012). Transferring the characteristics of naturally occurring and biased antibody repertoires to human antibody libraries by trapping CDRH3 sequences. PloS one *7*, e43471.

Victora, G.D., and Nussenzweig, M.C. (2012). Germinal centers. Annu. Rev. Immunol. *30*, 429-457.

Virnekas, B., Ge, L., Pluckthun, A., Schneider, K.C., Wellnhofer, G., and Moroney, S.E. (1994). Trinucleotide phosphoramidites: ideal reagents for the synthesis of mixed oligonucleotides for random mutagenesis. Nucleic Acids Res. *22*, 5600-5607.

Voigt, M., Braig, F., Gothel, M., Schulte, A., Lamszus, K., Bokemeyer, C., and Binder, M. (2012). Functional dissection of the epidermal growth factor receptor epitopes targeted by panitumumab and cetuximab. Neoplasia *14*, 1023-1031.

Wang, M.M. (2011). Notch signaling and Notch signaling modifiers. Int. J. Biochem. Cell Biol. *43*, 1550-1562.

Wang, Z., Longo, P.A., Tarrant, M.K., Kim, K., Head, S., Leahy, D.J., and Cole, P.A. (2011). Mechanistic insights into the activation of oncogenic forms of EGF receptor. Nat. Struct. Mol. Biol. *18*, 1388-1393.

Weiner, G.J. (2015). Building better monoclonal antibody-based therapeutics. Nat. Rev. Cancer *15*, 361-370.

Werkhoven, P.R., and Liskamp, R.M.J. (2013). Chemical approaches for localization, characterization and mimicry of peptide epitopes. In Biotherapeutics: recent developments using chemical and molecular biology, L.H. Jones, A.J. McKnight, eds. (Cambridge, UK: The Royal Society of Chemistry), pp. 263-284.

Wheeler, D.L., Dunn, E.F., and Harari, P.M. (2010). Understanding resistance to EGFR inhibitors-impact on future treatment strategies. Nat. Rev. Clin. Oncol. *7*, 493-507.

Whitehead, T.A., Chevalier, A., Song, Y., Dreyfus, C., Fleishman, S.J., De Mattos, C., Myers, C.A., Kamisetty, H., Blair, P., Wilson, I.A.*, et al.* (2012). Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. Nat. Biotechnol. *30*, 543-548.

Wu, Y., Cain-Hom, C., Choy, L., Hagenbeek, T.J., de Leon, G.P., Chen, Y., Finkle, D., Venook, R., Wu, X., Ridgway, J.*, et al.* (2010). Therapeutic antibody targeting of individual Notch receptors. Nature *464*, 1052-1057.

Yamashita-Kashima, Y., Iijima, S., Yorozu, K., Furugaki, K., Kurasawa, M., Ohta, M., and Fujimoto-Ouchi, K. (2011). Pertuzumab in combination with trastuzumab shows significantly enhanced antitumor activity in HER2-positive human gastric cancer xenograft models. Clin. Cancer Res.*17*, 5060-5070.

Yang, W., Yoon, A., Lee, S., Kim, S., Han, J., and Chung, J. (2017). Next-generation sequencing enables the discovery of more diverse positive clones from a phage-displayed antibody library. Exp. Mol. Med. *49*, e308.

Yarden, Y., and Sliwkowski, M.X. (2001). Untangling the ErbB signalling network. Nat. Rev. Mol. Cell. Biol. *2*, 127-137.

Yewale, C., Baradia, D., Vhora, I., Patil, S., and Misra, A. (2013). Epidermal growth factor receptor targeting in cancer: a review of trends and strategies. Biomaterials *34*, 8690-8707.

Zemlin, M., Klinger, M., Link, J., Zemlin, C., Bauer, K., Engler, J.A., Schroeder, H.W., Jr., and Kirkham, P.M. (2003). Expressed murine and human CDR-H3 intervals of equal length exhibit distinct repertoires that differ in their amino acid composition and predicted range of structures. J. Mol. Biol. *334*, 733-749.

Zhai, W., Glanville, J., Fuhrmann, M., Mei, L., Ni, I., Sundar, P.D., Van Blarcom, T., Abdiche, Y., Lindquist, K., Strohner, R.*, et al.* (2011). Synthetic antibodies designed on natural sequence landscapes. J. Mol. Biol. *412*, 55-71.

Zhang, H., Torkamani, A., Jones, T.M., Ruiz, D.I., Pons, J., and Lerner, R.A. (2011). Phenotype-information-phenotype cycle for deconvolution of combinatorial antibody libraries selected against complex systems. Proc. Natl. Acad. Sci. USA *108*, 13456-13461.

Zhang, Y., Liu, Y., Wang, Y., Schultz, P.G., and Wang, F. (2014). Rational design of humanized dual-agonist antibodies. J. Am. Chem. Soc.*137*, 38-41.

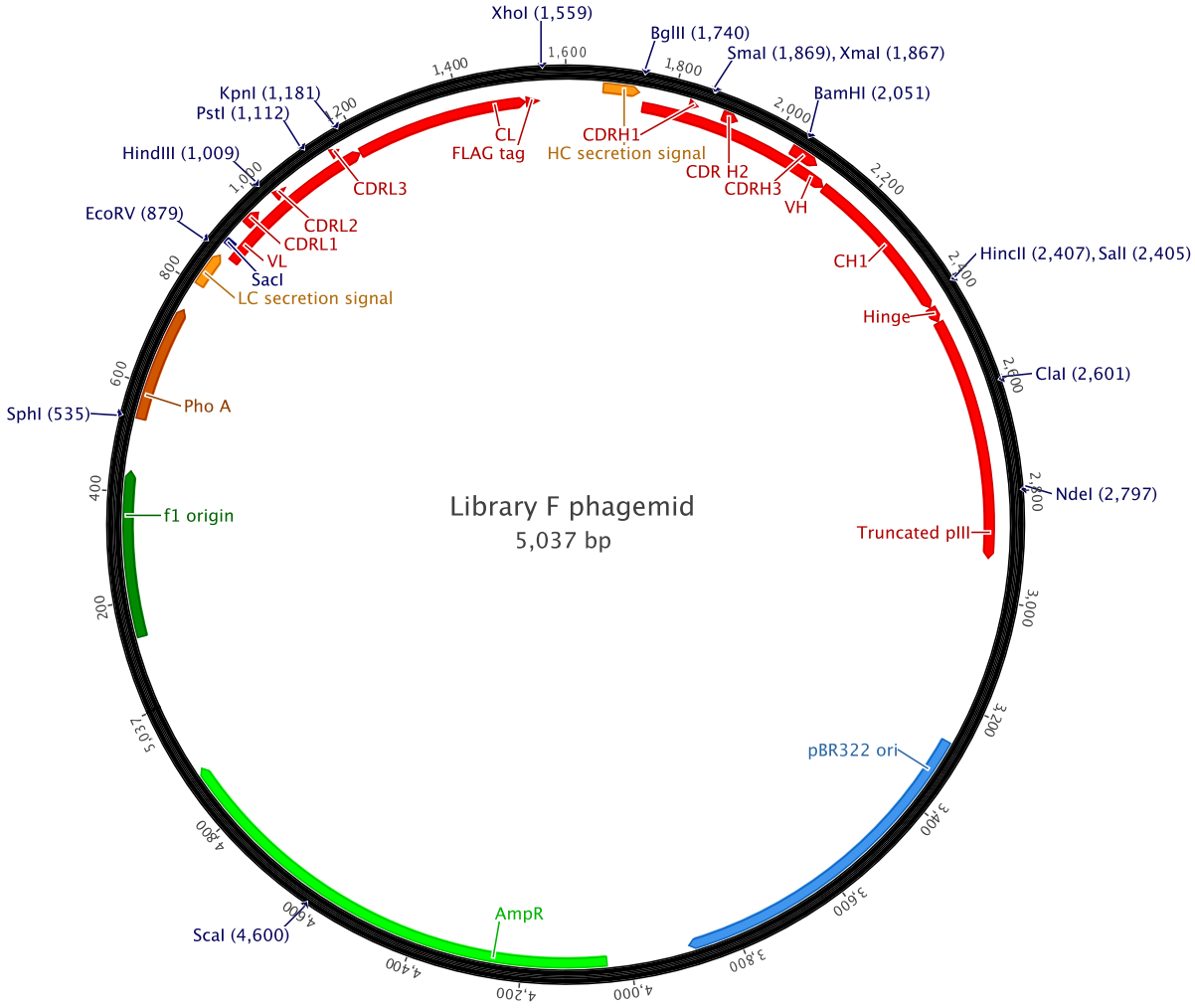
Zhang, Y., Wang, D., Welzel, G., Wang, Y., Schultz, P.G., and Wang, F. (2013). An antibody CDR3-erythropoietin fusion protein. ACS Chem. Biol. *8*, 2117-2121.

Zhang, Y., Zou, H., Wang, Y., Caballero, D., Gonzalez, J., Chao, E., Welzel, G., Shen, W., Wang, D., Schultz, P.G.*, et al.* (2015). Rational design of a humanized glucagon-like peptide-1 receptor agonist antibody. Angew. Chem. Int. Ed. Engl. *54*, 2126-2130.

Zhong, N., Loppnau, P., Seitova, A., Ravichandran, M., Fenner, M., Jain, H., Bhattacharya, A., Hutchinson, A., Paduch, M., Lu, V.*, et al.* (2015). Optimizing Production of Antigens and Fabs in the Context of Generating Recombinant Antibodies to Human Proteins. PLoS One *10*, e0139695.

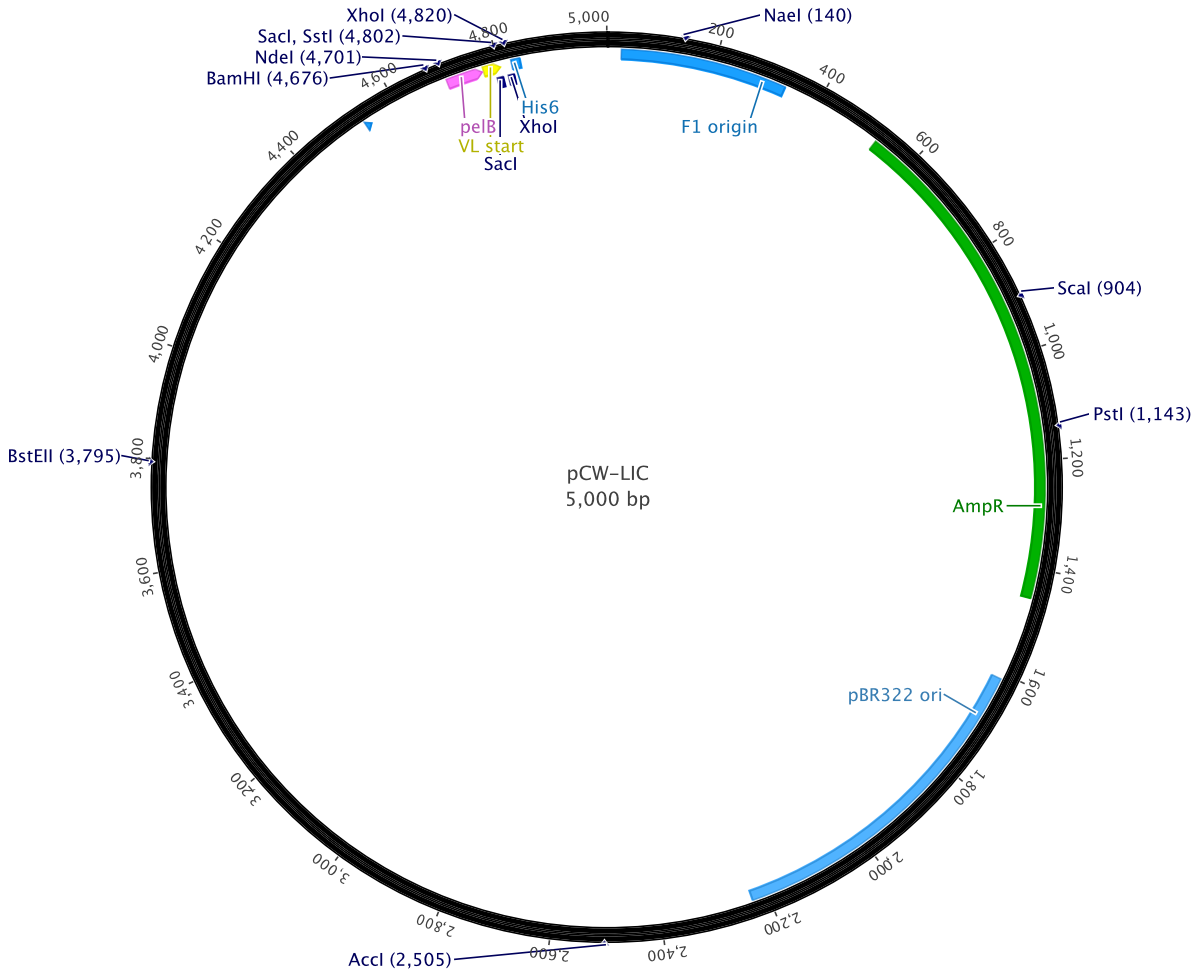
8. APPENDICES

Appendix 1: Library-F phagemid



The pHP-153 phagemid encoding the anti-MBP Fab is referred to as the library-F phagemid. It was used as a template phagemid for constructing library-F and the EGFR domain II structure-guided Fab library. It was also used for deriving template phagemids for constructing library-S and the modified-F library. The phagemid contains origins of replication for single-stranded DNA (f1 origin) and double-stranded DNA (pBR322 origin), and a selection marker (AmpR) that confers resistance to carbenicillin. An alkaline phosphatase-A (Pho-A) promoter drives the bicistronic expression of light and heavy chain fragments. The light chain contains the V_L domain, the C_L domain and a C-terminus FLAG tag. The heavy chain contains the V_H domain, the C_H1 domain, a hinge region and the truncated pIII protein. The N-terminus secretion signals direct the light and heavy chains to the bacterial periplasm where they associate to form Fabs.

Appendix 2: Fab expression plasmid



The pCW-LIC plasmid was used for expressing and purifying Fabs from *E. coli*. To simplify the sub-cloning of Fabs from the phagemid vector into the pCW-LIC plasmid, we modified the multiple cloning site of the pCW-LIC plasmid. After modification, the multiple cloning site contains a light chain secretion signal (pelB), eight amino acids from the V_L domain (V_L start), SacI and XhoI restriction sites, and a hexa-histidine tag. To sub-clone Fabs, Fab sequences were amplified from phagemids by PCR, and ligated into the SacI/XhoI-digested pCW-LIC vector using Gibson assembly. An IPTG-inducible Tac-promoter drives the bicistronic expression of light and heavy chain Fab fragments. The light chain contains the V_L domain, the C_L domain and a C-terminus FLAG tag. The heavy chain contains the V_H domain, the C_H1 domain, and a C-terminus hexa-histidine tag. The N-terminus secretion signals direct the light and heavy chains to the bacterial periplasm where they associate to form Fabs. The plasmid contains origins of replication for single-stranded DNA (f1 origin) and double-stranded DNA (pBR322 origin), and a selection marker (AmpR) that confers resistance to carbenicillin.

Appendix 3: List of oligonucleotides used for library-S CDR diversification

S-L3-09: TTCGCAACTTATTACTGTCAGCAA**ZZZZ**CCTCTGACGTTCGGACAGGGTACC

S-H3-JH4-07: CCGTCTATTATTGTGCTCGC**ZB**TTCGACTACTGGGGTCAAGGAAC
S-H3-JH4-08: CCGTCTATTATTGTGCTCGC**ZZB**TTCGACTACTGGGGTCAAGGAAC
S-H3-JH4-09: CCGTCTATTATTGTGCTCGC**ZZZB**TTCGACTACTGGGGTCAAGGAAC
S-H3-JH4-10: CCGTCTATTATTGTGCTCGC**ZZZZB**TTCGACTACTGGGGTCAAGGAAC
S-H3-JH4-11: CCGTCTATTATTGTGCTCGC**ZZZZZB**TTCGACTACTGG GGTCAAGGAAC
S-H3-JH4-12: CCGTCTATTATTGTGCTCGC**ZZZZZZB**TTCGACTACTGG GGTCAAGGAAC
S-H3-JH4-13: CCGTCTATTATTGTGCTCGC**ZZZZZZZB**TTCGACTACTGG GGTCAAGGAAC
S-H3-JH4-14: CCGTCTATTATTGTGCTCGC**ZZZZZZZZB**TTCGACTACTGG GGTCAAGGAAC
S-H3-JH4-15: CCGTCTATTATTGTGCTCGC**ZZZZZZZZZB**TTCGACTACTGGGGTCAAGGAAC
S-H3-JH4-16: CCGTCTATTATTGTGCTCGC**ZZZZZZZZZZB**TTCGACTACTGGGGTCAAGGAAC

S-H3-JH6-16: CCGTCTATTATTGTGCTCGC**XXXXXXX**TACTACTACTACTTTGACTACTGGGGTCAAGGAACCCT
S-H3-JH6-17: CCGTCTATTATTGTGCTCGC**XXXXXXXX**TACTACTAC**JO**GAC**U**TGGGGTCAAGGAACCCT
S-H3-JH6-18: CCGTCTATTATTGTGCTCGC**XXXXXXXXX**TACTACTAC**JO**GAC**U**TGGGGTCAAGGAACCCT
S-H3-JH6-19: CCGTCTATTATTGTGCTCGC**XXXXXXXXX**TACTACTACTAC**JO**GAC**U**TGGGGTCAAGGAACCCT
S-H3-JH6-20: CCGTCTATTATTGTGCTCGC**XXXXXXXXXX**TACTACTACTAC**JO**GACGTTTGGGGTCAAGGAACCCT
S-H3-JH6-21: CCGTCTATTATTGTGCTCGC**XXXXXXXXXXX**TACTACTACTAC**JO**GACGTTTGGGGTCAAGGAACCCT
S-H3-JH6-22: CCGTCTATTATTGTGCTCGC**XXXXXXXXXXXX**TACTACTACTAC**JO**GACGTTTGGGGTCAAGGAACCCT
S-H3-JH6-23: CCGTCTATTATTGTGCTCGC**XXXXXXXXXXXXX**TACTACTACTACGGA**O**GACGTTTGGGGTCAAGGAACCCT
S-H3-JH6-24: CCGTCTATTATTGTGCTCGC**XXXXXXXXXXXXXX**TACTACTACTACGGA**O**GACGTTTGGGGTCAAGGAACCCT
S-H3-JH6-25: CCGTCTATTATTGTGCTCGC**XXXXXXXXXXXXXXX**TACTACTACTACGGAATGGACGTTTGGGGTCAAGGAACCCT

21 mutagenic oligonucleotides were used for library-S diversification. Diversified positions within mutagenic oligonucleotides are colored in red. Codon Z denotes any of the following thirteen amino acids introduced at different proportions: Y (20%), S (20%), G (20%), T (6.5%), A (6.5%), P (6.5%), H (3.5%), R (3.5%), E (3.5%), F (2.5%), W (2.5%), V (2.5%) or L (2.5%). Codon B encodes for four amino acids A, G, D or Y at 25% each. Codon X denotes any of the following nine amino acids introduced at different proportions: Y (25%), S (20%), G (20%), A (10%), F (5%), W (5%), H (5%), P (5%) or V (5%). Codon J encodes for two amino acids G or Y at 50% each. Codon O encodes for two amino acids M or F at 50% each. Codon U encodes for two amino acids V or Y at 50% each.

Appendix 4: List of oligonucleotides used for diversifying CDRs in the modified-F library

F-L3-08: GCAACTTATTACTGTCAGCAA**XXXST**ACGTTCGGACAGGG
F-L3-09: GCAACTTATTACTGTCAGCAA**XXXXST**ACGTTCGGACAGGG
F-L3-10: GCAACTTATTACTGTCAGCAA**XXXXXST**ACGTTCGGACAGGG
F-L3-11: GCAACTTATTACTGTCAGCAA**XXXXXXST**ACGTTCGGACAGGG
F-L3-12: GCAACTTATTACTGTCAGCAA**XXXXXXXST**ACGTTCGGACAGGG

F-H1-13: GCAGCTTCTGGCTTCAAC**KWWWWQ**CACTGGGTGCGTCAG

F-H2-10: GCCTGGAATGGGTTGCA**W**ATT**WNWWRW**ACT**W**TATGCCGATAGCGTC

F-H3-JH4-07: CGTCTATTATTGTGCTCGC**XEF**TTCGACTACTGGGGTCAAG
F-H3-JH4-08: CGTCTATTATTGTGCTCGC**XXEF**TTCGACTACTGGGGTCAAG
F-H3-JH4-09: CGTCTATTATTGTGCTCGC**XXXEF**TTCGACTACTGGGGTCAAG
F-H3-JH4-10: CGTCTATTATTGTGCTCGC**XXXXEF**TTCGACTACTGGGGTCAAG
F-H3-JH4-11: CGTCTATTATTGTGCTCGC**XXXXXEF**TTCGACTACTGGGGTCAAG
F-H3-JH4-12: CGTCTATTATTGTGCTCGC**XXXXXXEF**TTCGACTACTGGGGTCAAG
F-H3-JH4-13: CGTCTATTATTGTGCTCGC**XXXXXXXEF**TTCGACTACTGGGGTCAAG
F-H3-JH4-14: CGTCTATTATTGTGCTCGC**XXXXXXXXEF**TTCGACTACTGGGGTCAAG
F-H3-JH4-15: CGTCTATTATTGTGCTCGC**XXXXXXXXXEF**TTCGACTACTGGGGTCAAG
F-H3-JH4-16: CGTCTATTATTGTGCTCGC**XXXXXXXXXXEF**TTCGACTACTGGGGTCAAG
F-H3-JH4-17: CGTCTATTATTGTGCTCGC**XXXXXXXXXXXEF**TTCGACTACTGGGGTCAAG
F-H3-JH4-18: CGTCTATTATTGTGCTCGC**XXXXXXXXXXXXEF**TTCGACTACTGGGGTCAAG
F-H3-JH4-19: CGTCTATTATTGTGCTCGC**XXXXXXXXXXXXXEF**TTCGACTACTGGGGTCAAG
F-H3-JH4-20: CGTCTATTATTGTGCTCGC**XXXXXXXXXXXXXXEF**TTCGACTACTGGGGTCAAG
F-H3-JH4-21: CGTCTATTATTGTGCTCGC**XXXXXXXXXXXXXXXEF**TTCGACTACTGGGGTCAAG
F-H3-JH4-22: CGTCTATTATTGTGCTCGC**XXXXXXXXXXXXXXXXEF**TTCGACTACTGGGGTCAAG
F-H3-JH4-23: CGTCTATTATTGTGCTCGC**XXXXXXXXXXXXXXXXXEF**TTCGACTACTGGGGTCAAG

S-H3-JH6-16: CCGTCTATTATTGTGCTCGC**XXXXXXX**TACTACTACTACTTTGACTACTGGGGTCAAGGAACCCT
S-H3-JH6-17: CCGTCTATTATTGTGCTCGC**XXXXXXXX**TACTACTAC**JO**GAC**U**TGGGGTCAAGGAACCCT
S-H3-JH6-18: CCGTCTATTATTGTGCTCGC**XXXXXXXXX**TACTACTAC**JO**GAC**U**TGGGGTCAAGGAACCCT
S-H3-JH6-19: CCGTCTATTATTGTGCTCGC**XXXXXXXXX**TACTACTACTAC**JO**GAC**U**TGGGGTCAAGGAACCCT
S-H3-JH6-20: CCGTCTATTATTGTGCTCGC**XXXXXXXXXX**TACTACTACTAC**JO**GACGTTTGGGGTCAAGGAACCCT
S-H3-JH6-21: CCGTCTATTATTGTGCTCGC**XXXXXXXXXXX**TACTACTACTAC**JO**GACGTTTGGGGTCAAGGAACCCT
S-H3-JH6-22: CCGTCTATTATTGTGCTCGC**XXXXXXXXXXXX**TACTACTACTAC**JO**GACGTTTGGGGTCAAGGAACCCT
S-H3-JH6-23: CCGTCTATTATTGTGCTCGC**XXXXXXXXXXXXX**TACTACTACTACGGA**O**GACGTTTGGGGTCAAGGAACCCT
S-H3-JH6-24: CCGTCTATTATTGTGCTCGC**XXXXXXXXXXXXXX**TACTACTACTACGGA**O**GACGTTTGGGGTCAAGGAACCCT
S-H3-JH6-25: CCGTCTATTATTGTGCTCGC**XXXXXXXXXXXXXXX**TACTACTACTACGGAATGGACGTTTGGGGTCAAGGAACCCT

34 mutagenic oligonucleotides were used for diversifying CDRs in the modified-F library. Diversified positions within mutagenic oligonucleotides are colored in red. Codon X denotes any of the following nine amino acids introduced at different proportions: Y (25%), S (20%), G (20%), A (10%), F (5%), W (5%), H (5%), P (5%) or V (5%). Codon S encodes for two amino acids P or L at 50% each. Codon T encodes for two amino acids I or F at 50% each. Codon K encodes for two amino acids I or L at 50% each. Codon W encodes for two amino acids Y or S at 50% each. Codon Q encodes for two amino acids I or M at 50% each. Codon N encodes for two amino acids P or S at 50% each. Codon R encodes for two amino acids G or S at 50% each. Codon E encodes for two amino acids A or G at 50% each. Codon F encodes for four amino acids F, L, I or M at 25% each. Codon J encodes for two amino acids G or Y at 50% each. Codon O encodes for two amino acids M or F at 50% each. Codon U encodes for two amino acids V or Y at 50% each.

Appendix 5: List of oligonucleotides used for diversifying CDRs in the EGFR domain II structure-guided Fab library

F-L3-08: GCAACTTATTACTGTCAGCAA**XXXST**ACGTTCGGACAGGG
F-L3-09: GCAACTTATTACTGTCAGCAA**XXXXST**ACGTTCGGACAGGG
F-L3-10: GCAACTTATTACTGTCAGCAA**XXXXXST**ACGTTCGGACAGGG
F-L3-11: GCAACTTATTACTGTCAGCAA**XXXXXXST**ACGTTCGGACAGGG
F-L3-12: GCAACTTATTACTGTCAGCAA**XXXXXXXST**ACGTTCGGACAGGG

F-H1-13: GCAGCTTCTGGCTTCAAC**KWWWWQ**CACTGGGTGCGTCAG

F-H2-10: GCCTGGAATGGGTTGCA**W**ATT**WNWWRW**ACT**W**TATGCCGATAGCGTC

EGFR-H3-38: TATTGTGCTCGCACT**D**TGTCCCCCCCTGATGCTGTACAACCCCACTACTTACCAAATGGACGTCAACCCCGAAGGTA
AATACTCTTTTGGTGCTACTTGT**D**TGGGCTATGGACTAC

Eight mutagenic oligonucleotides were used for diversifying CDRs in the EGFR domain II structure-guided Fab library. Diversified positions within mutagenic oligonucleotides are colored in red. Codon X denotes any of the following nine amino acids introduced at different proportions: Y (25%), S (20%), G (20%), A (10%), F (5%), W (5%), H (5%), P (5%) or V (5%). Codon S encodes for two amino acids P or L at 50% each. Codon T encodes for two amino acids I or F at 50% each. Codon K encodes for two amino acids I or L at 50% each. Codon W encodes for two amino acids Y or S at 50% each. Codon Q encodes for two amino acids I or M at 50% each. Codon N encodes for two amino acids P or S at 50% each. Codon R encodes for two amino acids G or S at 50% each. Codon D denotes the mixture of 15 amino acids encoded by the NNC codon.

Appendix 6: List of oligonucleotides used for NGS sample preparation

L3-Fwd: CCATCTCATCCCTGCGTGTCTCCGACTCAGAACCATCCGCCCGGAAGACTTCGCAACTTA

L3-Rev: CCTCTCTATGGGCAGTCGGTGATATCTCCACCTTGGTACCCTG

H1-Fwd: CCATCTCATCCCTGCGTGTCTCCGACTCAGAACCATCCGCCGTTTGTCCTGTGCAGCTTC

H1-Rev: CCTCTCTATGGGCAGTCGGTGATCCCTTACCCGGGGCCTGACG

H2-Fwd: CCATCTCATCCCTGCGTGTCTCCGACTCAGAACCATCCGCCCCCGGGTAAGGGCCTGGAA

H2-Rev: CCTCTCTATGGGCAGTCGGTGATCTTATAGTGAAACGGCCCTTGACGCT

H3-Fwd: CCATCTCATCCCTGCGTGTCTCCGACTCAGAACCATCCGCAGGACACTGCCGTCTATTAT

H3-Rev: CCTCTCTATGGGCAGTCGGTGATACGGTGACTAGTGTACCTTG

L3-H3 Seq: ACGTTCGGACAGGGTTATTATTGTGCTCGC

L3-H1 Seq: ACGTTCGGACAGGGTGCTTCTGGCTTCAAC

H1-H2 Seq: CACTGGGTGCGTCAGCTGGAATGGGTTGCA

H2-H3 Seq: TATGCCGATAGCGTCTATTATTGTGCTCGC

Oligonucleotides used for NGS amplicon preparation contained the following features: forward primers contained adaptor (blue), key (green), barcode (red) and antibody framework (black) regions, and reverse primers contained truncated P1 (orange) and antibody framework (black) regions. Adaptor and truncated P1 sequences were included in the primers for facilitating emulsion PCR following amplicon preparation. To amplify a CDR of interest from phage pools, for example CDRH3, H3-Fwd and H3-Rev primers were used in the PCR reaction. To generate the L3-H3 CDR strip from phage pools, one oligonucleotide (L3-H3 Seq) was used in the Kunkel mutagenesis reaction. To generate the L3-H1-H2-H3 CDR strip from phage pools, three oligonucleotides (L3-H1 Seq, H1-H2 Seq and H2-H3 Seq) were used in the Kunkel mutagenesis reaction. To amplify L3-H3 or L3-H1-H2-H3 CDR strips, L3-Fwd and H3-Rev primers were used in the PCR reaction.

Appendix 7: Contribution of Other Researchers to the PhD Project

Dr. Sachdev Sidhu at the University of Toronto provided library-F phage and library-F mutagenic oligonucleotides. Dr. Helena Persson (Sidhu Lab) designed and constructed library-F (Persson *et al*., 2013).

Wayne Hill: Sanger sequencing and Ion Torrent Sequencing

Daniel Hogan and Dr. Kris Barreto: Assistance with next-generation sequencing data analysis

Dr. Timothy Strozen: Modification of the pCW-LIC Fab expression plasmid

Dr. Landon Pastushok: Optimization of conditions for sub-cloning Fabs from the phagemid vector into the pCW-LIC Fab expression vector

The motif-grafting project described in Section 5.3 is a collaborative work between the Geyer lab at the University of Saskatchewan and the Sidhu lab at the University of Toronto. Dr. Geyer conceived the project, designed and constructed the EGFR domain II structure-guided library, and isolated the lead antibody fragment named DL06. Out of 8 figures in Section 5.3 (Figures 5.19 to 5.26), Dr. Shane Miersch (Sidhu Lab) generated data for Figures 5.20, 5.23A, 5.23C and 5.26C.

Lindsay Pelzer: Maintenance of HEK293F and A431 cell lines

Ashley Sutherland: Assistance with Adobe Illustrator